MODEL *COX PROPORTIONAL HAZARD* PADA KEJADIAN BERSAMA

SKRIPSI

Diajukan Kepada Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Negeri Yogyakarta
untuk Memenuhi Sebagian Persyaratan guna Memperoleh Gelar Sarjana Sains

Disusun Oleh:

Bayu M Iskandar
10305141018

PROGRAM STUDI MATEMATIKA
JURUSAN PENDIDIKAN MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI YOGYAKARTA
2015

PERSETUJUAN

Skripsi yang berjudul

"MODEL *COX PROPORTIONAL HAZARD* PADA KEJADIAN BERSAMA"

Oleh:

Bayu M Iskandar

NIM. 10305141018

Telah disetujui dan disahkan oleh dosen pembimbing untuk diujikan kepada Dewan Penguji Skripsi Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Yogyakarta

Disetujui pada tanggal:

22 Desember 2014

Menyetujui,

Dosen Pembimbing I	Dosen Pembimbing II
Rosita Kusumawati, M.Sc	Retno Subekti, M.Sc
NIP. 198007072005012001	NIP. 198111162005012002

PENGESAHAN

Skripsi yang berjudul

"MODEL *COX PROPORTIONAL HAZARD* PADA KEJADIAN BERSAMA"

disusun oleh:

Nama : Bayu M Iskandar
NIM : 10305141018
Prodi : Matematika

Skripsi ini telah diuji di depan Dewan Penguji Skripsi Fakulltas MIPA pada tanggal 2 Januari 2015 dan dinyatakan **Lulus**.

DEWAN PENGUJI

Nama	**Jabatan**	**Tanda Tangan**	**Tanggal**
Rosita Kusumawati, M.Sc **NIP. 19800707200501 2001**	Ketua Penguji	....	21-01-2015
Retno Subekti, M.Sc **NIP. 19811116200501 2002**	Sekretaris Penguji	....	20-01-2015
Mathilda Susanti, M.Si **NIP. 19640314198901 2001**	Penguji Utama	....	19-01-2015
Sahid, M.Sc **NIP. 19650905199101 1001**	Penguji Pendamping	....	14-01-2015

Yogyakarta, 21 Januari 2015
Fakultas Matematika dan Ilmu Pengetahuan Alam
Dekan,

Dr. Hartono
NIP. 196203291987021002

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini, saya:

Nama : Bayu M Iskandar

NIM : 10305141018

Program Studi : Matematika

Fakultas : Matematika dan Ilmu Pengetahuan Alam

Judul Skripsi : MODEL *COX PROPORTIONAL HAZARD* PADA KEJADIAN BERSAMA

Menyatakan bahwa skripsi ini benar-benar karya saya sendiri dan sepanjang pengetahuan saya, tidak terdapat karya atau pendapat yang ditulis atau diterbitkan orang lain, kecuali pada bagian-bagian tertentu yang diambil sebagai acuan atau kutipan dengan mengikuti tata penulisan karya ilmiah yang telah lazim.

Apabila ternyata terbukti pernyataan saya ini tidak benar, maka sepenuhnya menjadi tanggung jawab saya dan saya bersedia menerima sanksi sesuai ketentuan yang berlaku.

Yogyakarta, 22 Desember 2014
Yang Menyatakan,

Bayu Muchamad Iskandar
NIM 10305141018

MOTTO

"Kegagalan hanya terjadi bila kita menyerah."
(Lessing)

"Orang-orang yang sukses telah belajar membuat diri mereka melakukan hal yang harus dikerjakan ketika hal itu memang harus dikerjakan, entah suka atau tidak."
(Aldus Huxley)

"If you can dream it, you can do it"
(Walt E. Disney)

"Teman sejati adalah ia yang meraih tangan anda dan menyentuh hati anda"
(Heather Pryor)

"To get a success, your courage must be greater than your fear."
(Matthew)

"Hidup ini tak selalu berakhir indah, maka buatlah selalu indah sejak awal"
(Bayu M Iskandar)

HALAMAN PERSEMBAHAN

Kupersembahkan sebuah catatan kecil ini untuk:

- ♥ Malaikat Penjagaku, yang senantiasa membimbing dan mengajarkan aku menjadi yang terbaik. Malaikat yang tak pernah lelah untuk menyemangatiku dan selalu memberikan aku semua hal yang terindah. Terimakasih atas doamu disetiap sujudmu, terimakasih banyak malaikat duniaku, Bapak Sukandar Triwahadi dan Ibu Retno Utami.
- ♥ Kedua Adikku, alm. Tobby Takandara dan Vina Rosiana yang selalu menjadi semangat dan menghiburku setiap hari. Maaf kakak belum bisa menjadi contoh yang baik selama ini.
- ♥ Khoiruddin Aria Wijaya, Lina Febriani, Nanang Hermawan, Dimas Ridwan W, Ikfan Mida N, dan Doni Hermawan yang selalu menemani dalam pencarian jati diri saat kuliah dan selalu memberikan makna setiap saat.
- ♥ Sahabat hidupku yang akan terus selalu mengingatkanku, yang tanpa lelah memotivasiku dalam menghadapi semua masalah, terima kasih Anisa Jatus Anafauziah.
- ♥ Teman-teman Matsub 2010 semua, yang selalu kena kejailanku, yang selalu membantuku saat perkuliahan. Terimakasih banyak ya.
- ♥ Pihak-pihak yang membantuku dalam menempuh studi S1, yang tidak bisa saya tuliskan satu persatu, saya ucapkan terimakasih banyak.

MODEL *COX PROPORTIONAL HAZARD* PADA KEJADIAN BERSAMA

Oleh:
Bayu Muchamad Iskandar
10305141018

ABSTRAK

Model *Cox Proportional Hazard* pada kejadian bersama merupakan modifikasi dari model *Cox* ketika ada dua atau lebih individu yang mengalami kejadian bersama (*ties*). Penelitian ini bertujuan untuk menjelaskan prosedur pembentukan model *Cox* pada kasus kejadian bersama dan penerapannya untuk mengetahui faktor-faktor yang mempengaruhi kasus kecelakaan lalu lintas di Amerika Serikat.

Estimasi Parameter dalam prosedur pembentukan model *Cox* pada umumnya dengan menggunakan *Maximum Partial Likelihood Estimation* (MPLE) yaitu dengan memaksimalkan fungsi partial likelihood. Pada kasus kejadian bersama dilakukan modifikasi pada *partial likelihood* dengan pendekatan *Breslow*. Estimasi parameter dan perhitungan yang lainnya dalam penelitian ini dibantu dengan software R 3.0.3.

Prosedur pembentukan model *Cox* pada kejadian bersama terdiri dari 7 tahap, yaitu (1) identifikasi data, yaitu penentuan varibel-varibel yang akan digunakan dalam model cox. (2) Estimasi parameter dengan pendekatan metode *Breslow*. (3) Pemilihan model terbaik menggunakan *forward procedure.* (4) Pengujian parameter dengan menggunakan uji *wald.* (5) penyusunan model. (6) pengujian asumsi *proportional hazard* dan (7) interpretasi model. Data yang digunakan adalah 398 pengemudi yang pernah mengalami kecelakaan lalu lintas di Amerika Serikat pada 1 Januari 2010 yang diambil dari website *National Highway Traffic Safety Administration* (NHTSA) dengan 5 variabel bebas yaitu umur, jenis kelamin, kepemilikan surat ijin mengemudi (SIM), penggunaan sabuk pengaman, pengaruh alkohol. Penulis menggunakan variabel tersebut untuk mengetahui faktor-faktor yang mempengaruhi kecelakaan lalu lintas di Amerika Serikat. Berdasarkan pemilihan model terbaik diperoleh 3 variabel yang signifikan dalam model yaitu umur, kepemilikan SIM dan penggunaan sabuk pengaman. Selanjutnya dalam pengujian asumsi *proportional hazard* menunjukkan bahwa variabel SIM yang tidak memenuhi asumsi maka variabel tersebut dikeluarkan dari model. Sehingga diketahui faktor-faktor yang mempengaruhi kecelakaan lalu lintas di Amerika Serikat adalah umur pengemudi dan penggunaan sabuk pengaman.

Kata kunci: model *cox proportional hazard*, kejadian bersama, metode *breslow*, kecelakaan lalulintas.

KATA PENGANTAR

Puji syukur kehadirat Allah SWT sehingga penulis dapat menyelesaikan skripsi berjudul “MODEL *COX PROPORTIONAL HAZARD COX* PADA KEJADIAN BERSAMA”. Penulisan skripsi ini dibuat untuk memenuhi sebagian persyaratan guna memperoleh gelar Sarjana Sains Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Yogyakarta.

Skripsi ini tidak dapat diselesaikan tanpa bantuan, dukungan serta bimbingan beberapa pihak. Penulis mengucapkan terimakasih kepada:

1. Bapak Dr. Hartono selaku Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Yogyakarta yang telah memberikan kelancaran pelayanan dalam urusan akademik.
2. Bapak Dr. Sugiman selaku Ketua Jurusan Pendidikan Matematika Universitas Negeri Yogyakarta yang telah memberikan kelancaran pelayanan dalam urusan akademik.
3. Dr. Agus Maman Abadi, M.Si selaku Ketua Program Studi Matematika Universitas Negeri Yogyakarta serta Penasehat Akademik yang telah memberikan bimbingan serta motivasi selama studi.
4. Ibu Rosita Kusumawati, M.Sc dan Ibu Retno Subekti, M.Sc selaku dosen pembimbing yang telah berkenan memberikan waktu luang, arahan, bimbingan serta dengan penuh kesabaran meneliti setiap kata demi kata dalam skripsi ini.

5. Seluruh dosen Jurusan Pendidikan Matematika Universitas Negeri Yogyakarta yang telah memberikan ilmu kepada penulis.
6. Orangtua dan keluarga yang telah memberikan doa, dukungan, serta semangat kepada penulis.
7. Seluruh teman-teman matematika angkatan 2010 yang telah menghibur serta menyemangati penulis.
8. Semua pihak yang telah membantu penulisan skripsi ini hingga selesai.

Penulis menyadari adanya ketidaktelitian, kekurangan dan kesalahan dalam penulisan tugas akhir skripsi ini. Oleh karena itu, penulis menerima kritik dan saran yang bersifat membangun. Semoga penulisan tugas akhir ini dapat bermanfaat bagi pembaca dan pihak yang terkait.

Yogyakarta, 22 Desember 2014

Penulis

Bayu M Iskandar

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

BAB I

PENDAHULUAN

A. Latar Belakang Masalah

Analisis survival merupakan salah satu analisis dalam statistika, seperti halnya analisis runtun waktu. *Analisis survival* berbeda dengan analisis runtun waktu walaupun keduanya menggunakan waktu atau variabelnya bergantung pada waktu. Analisis runtun waktu menganalisis waktu periode dan sehingga dapat untuk menganalisis kejadian dimasa depan. Metode analisis runtun waktu yang digunakan dalam teknik peramalan seperti: *naive*, *simple averaging*, *moving average*, dan *autoregressive moving average* (ARMA) model. *Analisis survival* merupakan salah satu kumpulan dari prosedur statistika untuk analisis data dimana variabelnya adalah waktu sampai terjadinya suatu kejadian. Dalam *analisis survival*, terdapat variabel waktu sebagai waktu uji hidup (*survival time*), karena menunjukkan waktu itu seseorang telah bertahan (*survived*) selama periode tertentu. Demikian pula secara khusus menunjuk kejadian sebagai suatu kegagalan, karena kejadian yang menjadi perhatian biasanya adalah kematian, timbulnya penyakit atau beberapa kejadian lainnya (Kleinbaum & Klein, 2005: 4-5). Istilah *analisis survival* sering digunakan di bidang biostatistika dan kedokteran. Istilah ini juga dikenal dengan nama *event history analysis* di bidang ilmu sosiologi dan manajemen atau *failure-time analysis* di bidang *engineering*. Perbedaan antara analisis *survival* dengan prosedur analisis runtun waktu adalah konsep penyensoran. Data tersensor adalah data yang tidak bisa diamati secara utuh, karena adanya individu yang hilang ataupun dengan alasan lain, hingga tidak dapat diambil datanya atau sampai akhir pengamatan individu tersebut belum mengalami peristiwa tertentu. Jika berada dalam keadaan sebaliknya maka data tersebut disebut data tidak tersensor (Lee & Wang, 2003: 1-7).

Secara umum metode untuk mengestimasi dan kurva waktu *survival* dalam analisis *survival* yaitu metode tabel hidup (*Life Table), Actuarial (Cutler-Ederer), Cox Proportional Hazard* Model atau Model *Cox* dan metode *Product Limit (Kaplan Meier). Analisis survival* pada masa kini lebih banyak difokuskan pada fungsi hazard yaitu menganalisis peluang kejadian. Model *Cox* sering digunakan daripada metode lainnya karena dapat mengestimasi *hazard ratio* tanpa perlu diketahui fungsi *hazard* dasarnya, serta hasil dari model *Cox* hamper sama dengan hasil model parametrik. Waktu kejadian atau waktu *survival* dalam analisis *survival* terbagi menjadi 2 macam yaitu waktu kejadian tanpa ties dan waktu kejadian dengan *ties. Ties* atau kejadian bersama adalah keadaan yang terdapat dua individu atau lebih yang mengalami kejadian pada waktu yang bersamaan. Peneliti sering menghindari adanya data yang memiliki *ties* karena *ties* mengakibatkan permasalahan dalam membentuk *partial likelihoodnya* yaitu saat menentukan anggota dari himpunan risikonya.

Pendekatan untuk mengatasi kejadian bersama dalam analisis *survival* terdapat 3 metode yaitu metode *Efron*, metode *Breslow* dan metode *Exact* ((Breslow, 1974: 50). Metode *Breslow* merupakan metode yang sangat sederhana sehingga sering digunakan dalam mengatasi kejadian bersama. Metode *Breslow* mengasumsikan bahwa ukuran dari himpunan risiko untuk kejadian bersama adalah sama. Kesederhaan metode *Breslow* terdapat pada awal menganalisis karena tidak harus mengurutkan data atau kejadian mana yang terjadi lebih dahulu seperti pada metode lainnya.

Penelitian ini akan menggunakan model *Cox Proportional Hazard* dengan pendekatan metode *Breslow* pada kasus kecelakaan lalu lintas. Kecelakaan lalu lintas merupakan salah satu peristiwa yang banyak memakan korban jiwa. Tidak terkecuali untuk sebuah negara maju seperti Amerika Serikat dimana keselamatan sudah menjadi budaya yang diutamakan dalam aktivitas sehari-hari. Amerika Serikat merupakan negara dengan tingkat kepemilikan kendaraan tertinggi terjadi pada tahun 2007, yaitu 842,6 kendaraan per 1000 orang. Amerika Serikat memiliki jumlah kendaraan bermotor terbanyak di dunia, dengan jumlah 239,8 juta unit pada tahun 2010. Tingkat kepemilikan kendaraan per kapita di A.S. juga yang tertinggi di dunia, yaitu 769 kendaraan per 1000 orang, 10 mobil tiap 13 orang. Menurut Departemen Energi Amerika Serikat bahwa angka per kapita yang lebih tinggi, yaitu 828 kendaraan per 1000 penduduk dengan jumlah kendaraan 245,4 juta kendaraan. Amerika Serikat merupakan salah satu negara maju yang mempunyai peraturan berlalu lintas yang cukup lengkap dengan fasilitas pendukung yang memadai masih memiliki tingkat kematian akibat kecelakaan yang cukup tinggi. Hal tersebut dibuktikan dengan kejadian kecelakaan lalu lintas setiap tahunnya mencapai 500.000 kasus. Dari jumlah tersebut, 10% korban meninggal sebelum tiba di rumah sakit dan lebih dari 100.000 korban menderita berbagai tingkat kecacatan akibat kecelakaan lalu lintas tersebut. Berdasarkan penelitian, tingkat kecelakaan yang menelan korban jiwa kebanyakan dilakukan oleh anak muda dengan umur antara 16-25 tahun.

Pada hakikatnya penyebab kecelakaan tidak hanya berasal dari manusia akan tetapi ada beberapa elemen. Selama ini diketahui bahwa tiga elemen utama dari jalan raya yaitu manusia, kendaraan dan lingkungan (Haddon, 1980: 411-421). Ketiga elemen itu dimasukkan kedalam kerangka kerja keamanan jalan raya. Jika manajemen sistem hendak dilakukan secara efisien, maka seluruh faktor harus diperhitungkan. Faktor manusia adalah faktor yang mempengaruhi sebelum kecelakaan, pada saat kecelakaan dan setelah kecelakaan (Lay, 1986: 39). Sebagai contoh, pengemudi harus memiliki pengetahuan yang baik dalam berkendara dijalan raya. Selain itu dalam kejadian kecelakaan, kendaraan harus memiliki perangkat yang memadai untuk meminimalkan terjadinya korban jiwa, seperti sabuk keselamatan (*safety belt*), dan layanan medis darurat harus tersedia bagi korban kecelakaan. Lingkungan dan kendaraan juga harus memiliki atribut tertentu sebagai pencegahan atau menghindari kecelakaan.

Pada dasarnya pelanggaran lalu lintas yang bersifat sepele akan tetapi hal tersebut dapat mengakibatkan korban jiwa. Faktor manusia merupakan faktor yang paling mendominasi dalam peristiwa kecelakaan lalu lintas. Pelanggaran lalu lintas ini bisa terjadi karena kesengajaan melanggar lalu lintas, ketidaktahuan atau tidak adanya kesadaran terhadap arti aturan yang berlaku ataupun tidak melihat ketentuan yang berlaku dalam berkendaraan. Pelanggaran yang sering terjadi seperti tidak memakai helm, sabuk pengaman *(safety belt)*, dan sebagainya.

Kejadian kecelakaan lalu lintas dapat menimpa siapapun termasuk seseorang yang pernah mengalami kejadian di masa lalu. Terkadang pengemudi tidak belajar dari pengalamannya sendiri sehingga mengalami kecelakaan yang sama di masa datang. Kecelakaan juga dapat terjadi pada waktu bersamaan ditempat yang sama maupun berbeda disuatu wilayah.

Pada penelitian yang terkait dengan kecelakaan, Ahmad Wahidin (2008: 110) menganalisis tentang pengaruh penggunaan sabuk keselamatan terhadap tingkat fatalitas kecelakaan dan tingkat keparahan kecelakaan menggunakan metode uji *multivariate* (metode *multiple log regression*). Hasil penelitian tersebut menunjukkan bahwa sabuk pengaman berpengaruh signifikan pada tingkat fatalitas kecelakaan. Menurut Haryono Sukarto (1993) yang meneliti tentang interaksi faktor-faktor penyebab kecelakaan lalu lintas di jalan tol sekitar Jakarta dengan menggunakan analisis komponen utama (*principle component analysis*) menunjukkan bahwa faktor pengemudi penyebab kecelakaan yang paling besar pengaruhnya. Banyaknya kecelakaan yang diakibatkan oleh faktor-faktor seperti pengemudi dan sabuk pengaman, maka dengan *analisis survival* dapat dimodelkan suatu penyebab kecelakaan. Berdasarkan uraian diatas penulis ingin membahas model *cox propotional hazard* untuk memodelkan penyebab kecelakaan yang terjadi di wilayah Amerika Serikat.

B. Rumusan Masalah

Berdasarkan uraian latar belakang maka rumusan masalah yang akan dibahas dalam skripsi ini sebagai berikut.

1. Bagaimana prosedur pemodelan *cox proportional hazard* pada kejadian bersama?

2. Bagaimana penerapan model *cox proportional hazard* pada kejadian bersama dalam kasus kecelakaan lalu lintas di Amerika Serikat?

C. Tujuan Penulisan

Tujuan dari penulisan skripsi ini adalah sebagai berikut.

1. Menjelaskan prosedur pemodelan *cox proportional hazard* pada kasus kejadian bersama.
2. Hasil penerapan *cox proportional hazard* pada kasus kejadian bersama dalam kasus kecelakaan lalu lintas di Amerika Serikat.

D. Manfaat Penulisan

Penulisan skripsi ini diharapkan dapat memberikan manfaat sebagai berikut.

1. Bagi mahasiswa

   Pengembangan ilmu teoritis yang dipelajari diperkuliahan dan penambahan wawasan *Analisis survival* khususnya metode *proportional hazard model* dengan variabel risiko yang bergantung terhadap waktu.

2. Bagi penulis

   Menambah pengetahuan mengenai *Analisis survival* khususnya model *cox proportional hazard* dengan variabel risiko yang bergantung terhadap waktu.

3. Bagi Perpustakaan Jurusan Pendidikan Matematika UNY

   Menambah referensi mengenai model *cox proportional hazard* dengan variabel risiko yang bergantung terhadap waktu bagi mahasiswa Jurusan Pendidikan Matematika.

4. Bagi Pembaca

   Penelitian ini dapat dijadikan referensi pembaca untuk mengembangkan model *cox proportional hazard* pada kasus-kasus yang terjadi di Indonesia.

BAB II

KAJIAN PUSTAKA

Pada Bab II akan dibahas tentang kecelakaan, *analisis survival*, fungsi kepadatan peluang, fungsi *survival*, fungsi peluang bersyarat, kejadian bebas atau *independent,* fungsi *hazard, hazard kumulatif,* model *cox proportional hazard, residual model cox propotional hazard,* pengujian asumsi *proportional hazard,* serta interprestasi model *cox proportional hazard*

A. Kecelakaan

1. Definisi Kecelakaan

Berdasarkan Undang-Undang Republik Indonesia nomor 22 tahun 2009 tentang Lalu Lintas dan Angkutan Jalan, pasal 1 ayat (24) kecelakaan lalu lintas adalah suatu peristiwa di jalan yang tidak di sangka-sangka dan tidak disengaja melibatkan kendaraan dengan atau pemakai jalan lainnya, mengakibatkan korban manusia atau kerugian harta benda.

2. Faktor-Faktor Kecelakaan Lalu Lintas

Empat faktor utama yang menyebabkan terjadinya kecelakaan lalu lintas (Rinto, 2014: 15).

a. Faktor Manusia

Faktor manusia merupakan faktor yang paling dominan dalam kecelakaan. Hampir semua kejadian kecelakaan didahului dengan pelanggaran rambu-rambu lalu lintas. Pelanggaran dapat terjadi karena sengaja melanggar, ketidaktahuan terhadap arti aturan yang berlaku ataupun tidak melihat ketentuan yang di berlakukan atau pura-pura tidak tahu. Selain itu manusia sebagai pengguna jalan raya sering lalai bahkan ugal-ugalan dalam mengendarai kendaraan, tidak sedikit angka kecelakaan lalu lintas diakibatkan karena mengendarai kendaraan dalam pengaruh alkohol (mabuk), mengantuk, dan belum memiliki pengetahuan lebih dalam berpengendara (tidak memiliki SIM).

b. Faktor Kendaraan

Faktor kendaraan yang paling sering adalah kelalaian perawatan yang dilakukan terhadap kendaraan. Hal yang ssering terjadi pada saat kecelakaan seperti ban pecah, rem tidak berfungsi sebagaimana seharusnya, kelelahan logam yang mengakibatkan bagian kendaraan patah, peralatan yang sudah aus tidak diganti dan lain-lainnya. Untuk mengurangi faktor kendaraan, perawatan dan perbaikkan kendaraan diperlukan, disamping itu adanya kewajiban untuk melakukan pengujian kendaraan bermotor secara teratur.

c. Faktor Jalan

Faktor jalan terkait dengan kecepatan kendaraan, geometrik jalan, pagar pengaman di daerah pegunungan, ada tidaknya median jalan, jarak pandang dan kondisi permukaan jalan. Jalan yang rusak atau berlubang sangat pembahayakan semua pemakai jalan terutama pemakai sepeda motor.

d. Faktor Cuaca atau Lingkungan

Faktor cuaca seperti hujan juga mempengaruhi kerja kendaraan seperti jarak pengereman menjadi lebih jauh, jalan menjadi lebih licin, jarak pandang juga terpangaruh karena penghapus kaca tidak bisa bekerja secara sempurna atau lebatnya hujan mengakibatkan jarak pandang menjadi pendek. Asap dan kabut juga bisa mengganggu jarak pandang, terutama di daerah pegunungan.

B. Analisis *Survival*

Analisis *survival* telah menjadi alat penting untuk menganalisis data waktu antar kejadian (*time to event data*) atau menganalisis data yang berhubungan dengan waktu, mulai dari *time origin* sampai terjadinya suatu peristiwa khusus. Kejadian khusus (*failure event)* tersebut dapat berupa kegagalan, kematian, kambuhnya suatu penyakit, respon dari suatu percobaan, atau peristiwa lain yang dipilih sesuai dengan kepentingan peneliti. Peristiwa khusus tersebut dapat berupa kejadian positif seperti kelahiran, kelulusan sekolah, kesembuhan dari suatu penyakit (Kleinbaum & Klein, 2005: 4).

Analisis *survival* banyak diterapkan dalam bidang biologi, kedokteran, kesehatan umum seperti daya hidup pasien kanker paru-paru, sosiologi, teknik, seperti menganalisis masa hidup lampu pijar, ekonomi, demografi, dan epidemilogi (Collett, 2003: 1).

1. **Data *Survival***

Data *survival* merupakan data tentang pengamatan jangka waktu dari awal pengamatan sampai terjadinya suatu peristiwa. Waktu *survival* dapat didefinisikan sebagai waktu dari awal pengamatan hingga terjadinya peristiwa gagal, dapat dalam hari bulan, maupun tahun. Waktu awal (*time origin* atau *start-point*) yaitu waktu pada saat terjadinya kejadian awal, seperti waktu seorang divonis menderita kanker, waktu pemberian perlakuan dan lain-lain. Waktu kegagalan (*failure time* atau *end-point*) yaitu waktu pada saat terjadinya kejadian akhir seperti kematian, kejadian dan lain-lain (Collett, 2003: 1).

Penentuan waktu *survival*, ada tiga faktor yang dibutuhkan.

a. Waktu awal pencatatan (*time origin* atau *start-point*) harus didefinisikan dengan tepat pada setiap individu, misalkan awal mula pengamatan berupa tanggal perawatan pasien.

b. Waktu akhir pencatatan (*failure time* atau *end-point*) didefinisikan jelas untuk mengetahui status tersensor atau tidak tersensor, meninggal atau sembuh seorang pasien.

c. Skala pengukuran sebagai batas dari waktu kejadian dari awal sampai akhir kejadian, misalnya skala tahunan, bulanan, harian, mingguan, harian.

Data tersensor merupakan data yang tidak bisa diamati secara utuh, karena adanya individu yang hilang ataupun dengan alasan lain, sehingga tidak dapat diambil datanya sampai akhir pengamatan. Dengan kata lain, pada akhir pengamatan individu tersebut belum mengalami peristiwa tertentu. Jika berada dalam keadaan sebaliknya maka data tersebut disebut data tidak tersensor (Lee & Wang, 2003: 2).

2. **Tipe Penyensoran**

Dalam mendapatkan data *survival* sering dijumpai suatu individu tidak mengalami kejadian sampai batas waktu pengamatan. Biasanya untuk mendapatkan data *survival* yang lengkap sampai semua individu mengalami kejadian membutuhkan waktu yang lama sehingga pengamatan yang dilakukan tidak efektif dan mengakibatkan biaya yang dikeluarkan sangat banyak. Untuk mengatasi hal tersebut maka perlu dilakukan pensensoran data. Konsep penyensoran inilah yang membedakan antara analisis *survival* dengan ilmu-ilmu statistika yang lainnya (Kleinbaum & Klein, 2005: 5).

Menurut Klein & Moeschberger (2003: 64-70) dalam analisis *survival* terdapat empat jenis penyensoran.

a. Penyensoran kanan (*right censoring*)

   Penyensoran terjadi jika objek pengamatan atau individu yang diamati masih tetap hidup pada saat waktu yang telah ditentukan. Dengan kata lain individu tersebut belum mengalami kejadian sampai akhir periode pengamatan, sedangkan waktu awal dari objek pengamatan dapat diamati secara penuh. Sebagai contoh, seorang pasien kanker diamati dari awal perawatan sampai akhir perawatan ternyata pasien tersebut masih hidup. Kemudian pasien melanjutkan perawatan di luar negeri sehingga tidak bisa diamati lagi (*lost to follow up*). Pasien ini memiliki waktu *survival* setidaknya beberapa waktu. Sehingga waktu pengamatan individu tersebut dikatakan penyensoran kanan.

b. Penyensoran kiri (*left censoring*)

   Penyensoran kiri terjadi jika semua informasi yang diinginkan diketahui dari seseorang individu telah diperoleh pada awal pengamatan. Dengan kata lain pada saat waktu awal pengamatan individu tidak teramati pada awal pengamatan sementara kejadian dapat diamati secara penuh sebelum penelitian berakhir. Sebagai contoh, dalam sebuah penelitian untuk menentukan sebaran pengguna ganja di kalangan anak laki-laki di sebuah sekolahan. Dengan mengajukan pertanyaan "kapan pertama kali anda menggunakan ganja?". Ternyata terdapat beberapa anak menjawab "saya perrnah menggunakannya, tetapi saya tidak tahu tepatnya kapan pertama kali mengunakannya", pada kasus ini anak tersebut mengalami penyensoran kiri.

c. Penyensoran selang (*interval censoring*)

   Penyensoran selang terjadi jika informasi yang dibutuhkan telah dapat diketahui pada kejadian peristiwa di dalam selang pengamatan atau penyensoran yang waktu daya tahannya berada dalam suatu selang tertentu. Sebagai contoh, beberapa tikus yang diberikan karsinogen pada makanannya, dilakukan studi selama 10 bulan kepada 10 tikus dan penelitian dilakukan setiap akhir tahun, jika 2 dari 8 tikus tewas karena kanker pada bulan ke-5 dan ke-7, maka dua tikus tersebut mengalami penyensoran selang.

d. Penyensoran acak (*random censoring*)

Penyensoran acak terjadi jika individu yang diamati meninggal atau mengalami kejadian karena sebab yang lain, bukan disebabkan dari tujuan utama penelitian. Sebagai contoh, 10 tikus yang diberikan zat karsinogen pada makanannya. Pada saat pengamatan ada 1 dari 10 tikus tersebut meninggal karena terjepit (tewas bukan karena penelitian utama) bukan karena terkena kanker, maka tikus tersebut mengalami pensensoran acak.

Penyensoran-penyensoran di atas disebabkan oleh beberapa hal antara lain: (Kleinbaum & Klein, 2005: 6).

a. *Loss to follow up*, objek menghilang selama masa pengamatan terjadi apabila individu pindah atau menolak untuk berpartisipasi.
b. Individu tidak mengalami kejadian gagal (*failure event*) sebelum pengamatan berakhir.
c. Individu terpaksa dihentikan dari pengamatan karena kematian (jika kematian bukan failure event) atau disebabkan alasan lain.

C. Fungsi Kepadatan Peluang

Fungsi kepadatan peluang adalah peluang suatu individu mati atau gagal dalam interval waktu t sampai $t + \Delta t$. Fungsi kepadatan peluang dinotasikan dengan $f(t)$ dan dirumuskan dengan

$$f(t) = \lim_{\Delta t \to 0} \left[\frac{P\big(t < T < (t + \Delta t)\big)}{\Delta t} \right] = \lim_{\Delta t \to 0} \left[\frac{P\big(F(t + \Delta t) - F(t)\big)}{\Delta t} \right]. \qquad (2.1)$$

Misalkan T adalah variabel random bukan negatif pada interval $[0, \infty)$ yang menunjukkan waktu hidup pada suatu populasi dan $f(t)$ merupakan fungsi kepadatan peluang dari s maka fungsi distribusi kumulatif $F(t)$ adalah (Lawless, 1982:8)

$$F(t) = P(T \leq t)$$

$$= \int_0^t f(x)dx. \qquad (2.2)$$

Dari persamaan (2.2), diperoleh

$$f(t) = \frac{d\big(F(t)\big)}{dt} = F'(t). \qquad (2.3)$$

D. Fungsi *Survival*

Jika T merupakan variabel random tidak negatif pada interval $[0, \infty)$ yang menunjukkan waktu individu sampai mengalami kejadian pada populasi, $f(t)$ merupakan fungsi kepadatan peluang dari t maka peluang suatu individu tidak mengalami kejadian sampai waktu t dinyatakan dengan fungsi *survival* $S(t)$ (Lawless, 2007: 8).

$$S(t) = P(T \geq t)$$

$$= \int_t^\infty f(x)dx. \qquad (2.4)$$

Dari definisi fungsi distribusi kumulatif dari T, fungsi *survival* dapat dinyatakan sebagai berikut.

$$S(t) = P(T \geq t)$$

$$= 1 - P(T \leq t)$$

$$= 1 - F(t)$$

$$F(t) = 1 - S(t)$$

$$\frac{d\big(F(t)\big)}{dt} = \frac{d\big(1 - S(t)\big)}{dt}$$

$$f(t) = -\frac{d\big(S(t)\big)}{dt} = -S'(t). \tag{2.5}$$

Hubungan kepadatan peluang, fungsi distribusi kumulatif dari T dengan fungsi *survival* yaitu

$$f(t) = F'(t). = -S'(t). \tag{2.6}$$

E. Peluang Bersyarat

Peluang kejadian H dengan syarat K dinyatakan dengan lambang $P(H|K)$.

$$P(H|K) = \frac{P(H \cap K)}{P(K)} \; untuk \; P(K) > 0. \tag{2.7}$$

F. Kejadian Bebas atau *Independen*

Pada umumnya $P(A|B) \neq P(A)$ yang berarti terjadinya A dipenuhi oleh terjadinya B, namun bila $P(A|B) = P(A)$ berarti terjadinya B tidak mempengaruhi terjadinya A, maka A dan B disebut independen. Pernyataan dapat disimbolkan sebagai berikut.

Dua kejadian A dan B disebut independen apabila sebagai berikut.

a. $P(A|B) = P(A)$.

b. $P(B|A) = P(B)$.

c. $P(A \cap B) = P(A)P(B)$.

G. Fungsi *Hazard*

Misalkan T variabel random non negatif pada interval $[0, \infty)$ yang menunjukkan waktu individu sampai mengalami kejadian pada suatu populasi, maka peluang bahwa individu mengalami kejadian pada interval $(t, t + \Delta t)$ dinyatakan dengan fungsi *hazard* $h(t)$ (Lawless, 2007: 8).

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < \ t + \Delta t |\, T \geq t)}{\Delta t}$$

$$= \lim_{\Delta t \to 0} \frac{P(t \leq T < \ t + \Delta t,\ T \geq t)}{\Delta t. P(T \geq t)}$$

$$= \lim_{\Delta t \to 0} \frac{P(t \leq T < \ t + \Delta t)}{\Delta t. S(t)}$$

$$= \lim_{\Delta t \to 0} \frac{F(t + \Delta t) - F(t)}{\Delta t. S(t)}$$

$$= \frac{1}{S(t)} \lim_{\Delta s \to 0} \frac{F(t + \Delta t) - F(t)}{\Delta t}$$

$$= \frac{F'(t)}{S(t)}$$

$$= \frac{f(t)}{S(t)}. \tag{2.8}$$

H. *Hazard* Kumulatif

Dari hasil substitusi persamaan (2.5) dan (2.8) diperoleh sebagai berikut (Lawless, 2007)

$$h(t) = \frac{S'(t)}{S(t)} = \frac{d}{dt}\log S(t). \tag{2.9}$$

Berdasarkan persamaan (2.10) diperoleh (Lawless, 2007)

$$\log\big(S(t)\big)|_0^t = -\int_0^t h(x)\,dx \tag{2.10}$$

karena $S(0) = 1$ sehingga (Lawless, 2007)

$$S(t) = \exp\left[-\int_0^t h(x)\,dx\right]. \tag{2.11}$$

Dari persamaaan (2.11) didapatkan fungsi *hazard* maka fungsi kumulatif *hazard* dinyatakan dengan $H(t)$ (Lawless, 2007: 9).

$$H(t) = \int_0^t h(x)dx. \tag{2.12}$$

Selain itu persamaan (2.11) dapat dituliskan (Lawless, 2007).

$$S(t) = \exp[-H(t)]. \qquad (2.13)$$

I. Model *Cox Propotional Hazard*

Model *cox proportional hazard* disebut dengan model *cox* karena asumsi *proportional hazardnya* yaitu fungsi *hazard* dari individu yang berbeda adalah *proportional* atau rasio dari fungsi *hazard* dua individu yang berbeda adalah konstan (Lee & Wang, 2003: 298). Model *Cox* merupakan model berdistribusi semiparametrik karena dalam model *Cox* tidak memperlukan informasi tentang distribusi yang mendasari waktu *survival* dan untuk mengestimasi parameter regresi dari model *Cox* tanpa harus menentukan fungsi *hazard* dasar (Guo, 2009: 73).

Melalui model *Cox* dapat dilihat hubungan antara variabel bebas (variabel independen) terhadap variabel terikat (variabel dependen) yaitu waktu *survival* melalui fungsi *hazardnya*. Risiko kematian individu pada waktu tertentu bergantung pada nilai $x_1, x_2, \ldots, x_p$ dari p variabel bebas $X_1, X_2, \ldots, X_p$. Himpunan nilai variabel bebas pada model *Cox* dipresentasikan oleh vektor $\boldsymbol{x}$, sehingga $\boldsymbol{x} = \left(x_1, x_2, \ldots, x_p\right)$. Diasumsikan X merupakan variabel bebas yang independen terhadap waktu. Model Cox dapat dituliskan sebagai berikut.

$$h(t, x) = h_0(t) exp\left(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p\right) \qquad (2.14)$$

dengan memisalkan,

$h_0(t)$ = fungsi dasar *hazard*,

$\beta_1, \beta_2, \ldots, \beta_p$ = parameter regresi,

$x_1, x_2, \dots, x_p$ = nilai dari variabel bebas $X_1, X_2, \dots, X_p$.

Rumus model Cox pada persamaan (2.14) memiliki sifat bahwa jika semua X sama dengan nol, maka rumus tereduksi menjadi fungsi *hazard* dasar $h_0(t)$. Dengan demikian $h_0(t)$ dianggap sebagai awal atau dasar dari fungsi *hazard*, dapat dituliskan sebagai berikut.

$$\begin{aligned} h(t,x) &= h_0(t)exp\left(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p\right) \\ &= h_0(t)exp\left(\beta_1 \times 0 + \beta_2 \times 0 + \cdots + \beta_p \times 0\right) \\ &= h_0(t)exp(0) \\ &= h_0(t)(1) \\ h(t,x) &= h_0(t). \end{aligned} \tag{2.15}$$

Persamaan (2.14) dapat dituliskan dalam bentuk sebagai berikut.

$$\log h(t,x) = \log[h_0(t)] + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p$$

$$\log \frac{h(t,x)}{h_0(t)} = \left(\beta_1 x_1 + \beta_2 x 2_2 + \cdots + \beta_p x_p\right). \tag{2.16}$$

Model Cox mengestimasi parameter regresi $(\beta_1, \beta_2, \dots, \beta_p)$ dilakukan tanpa mengestimasi fungsi *hazard* dasar. Model pada persamaan (2.15) merupakan model dari log *hazard* rasio. Rasio *hazard* didefinisikan sebagai *hazard* dari satu individu dibagi dengan *hazard* individu yang berbeda (Kleinbaum & Klein, 2005). Persamaan (2.15) dapat dinyatakan sebagai berikut.

$$\log[HR(x)] = \left(\beta_1 x_1 + \beta_{22} x_2 + \cdots + \beta_p x_p\right). \tag{2.17}$$

Persamaan (2.17) mengimplikasikan bahwa dalam model dengan variabel bebas $x_1, x_2, \dots, x_p$ dan koefisien β_j yaitu peningkatan pada log rasio *hazard* untuk peningkatan satu satuan variabel bebas x_j , dengan asumsi bahwa nilai dari variabel bebas yang lain konstan. Dengan kata lain $exp(\beta_1)$ adalah rasio *hazard* untuk peningkatan satu satuan dalam x_j . Ketika variabel bebas dengan rasio *hazard* kurang dari $1(\ \beta > 0)$, peningkatan nilai variabel bebas berhubungan dengan lebih menurunnya risiko dan lebih panjangnya waktu bertahan hidup. Ketika rasio *hazard* lebih besar dari $1(\ \beta > 0)$, peningkatan nilai variabel bebas berhubungan dengan peningkatan risiko dan lebih pendeknya waktu bertahan hidup (Vittinghoff, Glidden, Shiboski, & McCulloch, 2004: 207).

1. Estimasi Parameter

Parameter β_j pada model *cox proportional hazard* akan diestimasi dengan menggunakan metode *Maximum Partial Likelihood Estimation* (MPLE). Pendugaan β_j dengan metode MPLE adalah nilai ketika fungsi partial likelihood maximum. Misal data untuk n individu yang terdiri dari r waktu kejadian yang tidak tersensor dan n-r individu tersensor kanan, diurutkan menjadi $t_1 < t_2 < \cdots < t_j \ldots < t_n$ dengan t_j merupakan urutan waktu kejadian ke-j .

Diasumsikan hanya terdapat satu individu yang mengalami kematian pada tiap waktu kegagalan, jadi tidak terjadi *ties* pada data. *Ties* adalah keadaan dimana terdapat dua individu atau lebih yang mengalami kejadian gagal pada waktu yang sama. Hal lain, yang perlu dipertimbangkan adalah peluang kematian suatu individu yang mati pada waktu kegagalan t_j, dengan syarat t_j menjadi salah satu yang diamati dari r waktu kegagalan $t_1, t_2, \dots, t_r$. Jika vektor variabel bebas dari individu yang mati pada waktu t_j ,dinotasikan dengan x_j , maka peluangnya menjadi sebagai berikut.

P[individu dengan variabel x_j mati pada t_j| satu kematian pada t_j].

Seperti pada persamaan (2.7), misalkan kejadian A adalah individu dengan variabel x_j meninggal pada saat t_j dan kejadian B adalah semua kematian pada saat t_j , maka

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)}$$

$$= \frac{P[\text{individu dengan variabel } x_j \text{ mati pada } t_j]}{P[\text{semua kematian pada } t_j]}. \qquad (2.18)$$

Pembilang pada persamaan (2.18) adalah bentuk sederhana dari risiko kematian pada waktu t_j untuk individu dengan variabel x_j . Jika pembilang tersebut adalah individu ke-i yang meninggal pada saat t_j , fungsi *hazard* ini dapat ditulis menjadi $h_i(t_j)$. Penyebutnya adalah penjumlah dari peluang kematian pada waktu t_j (dinotasikan $h_i(t_j)$) dari semua individu yang mempunyai risiko kematian pada waktu t_j . Dengan R(t_j) adalah himpunan individu yang berisiko pada waktu t_j yang terdiri dari individu-individu yang bertahan hidup hingga t_j .

Sehingga peluang dalam persamaan (2.17) menjadi $\frac{h_i(t_j)}{\sum_{i\in \mathrm{R}(t_j)} h_i(t_j)}$, menggunakan persamaan (2.14), maka fungsi *hazard* dasar adalah

$$P(A|B) = \frac{h_i(t_j)}{\sum_{i\in \mathrm{R}(t_j)} h_i(t_j)}$$

$$= \frac{h_0(t)\exp\left(\sum_{j=1}^{p}\beta_j x_{ij}\right)}{\sum_{l\epsilon R(t_j)} h_0(t)\exp\left(\sum_{j=1}^{p}\beta_j x_{lj}\right)}$$

$$= \frac{h_0(t)\exp\left(\sum_{j=1}^{p}\beta_j x_{ij}\right)}{h_0(t)\sum_{l\epsilon R(t_j)}\exp\left(\sum_{j=1}^{p}\beta_j x_{lj}\right)}$$

$$= \frac{\exp\left(\sum_{j=1}^{p}\beta_j x_{ij}\right)}{\sum_{l\epsilon R(t_j)}\exp\left(\sum_{j=1}^{p}\beta_j x_{lj}\right)}. \tag{2.19}$$

Dengan mengambil hasil peluang bersyarat diatas, memberikan fungsi partial likelihood sebagai berikut

$$L(\beta) = \prod_{i=1}^{r}\frac{\exp\left(\sum_{j=1}^{p}\beta_j x_{ij}\right)}{\sum_{l\epsilon R(t_j)}\exp\left(\sum_{j=1}^{p}\beta_j x_{lj}\right)}. \tag{2.20}$$

Dari persamaan (2.20) diperoleh fungsi *log partial likelihood* yaitu sebagai berikut.

$$\ln L(\beta) = \ln\prod_{i=1}^{r}\frac{\exp\left(\sum_{j=1}^{p}\beta_j x_{ij}\right)}{\sum_{l\epsilon R(t_j)}\exp\left(\sum_{j=1}^{p}\beta_j x_{lj}\right)}$$

$$= \sum_{i=1}^{n} \left[\ln\left(\exp\left(\sum_{j=1}^{p} \beta_j x_{ij} \right) \right) - \ln\left(\sum_{l\epsilon R(t_j)} \exp\left(\sum_{j=1}^{p} \beta_j x_{lj} \right) \right) \right]$$

$$= \sum_{i=1}^{n} \left[\left(\sum_{j=1}^{p} \beta_j x_{ij} \right) - \ln\left(\sum_{l\epsilon R(t_j)} \exp\left(\sum_{j=1}^{p} \beta_j x_{lj} \right) \right) \right]. \tag{2.21}$$

Turunan pertama dari $\ln L(\beta)$ terhadap β_j yaitu sebagai berikut

$$\frac{\partial \ln L(\beta)}{\partial \beta_j} = \frac{\partial \left(\sum_{i=1}^{n} \left[\left(\sum_{j=1}^{p} \beta_j x_{ij} \right) - \ln\left(\sum_{l\epsilon R(t_j)} \exp\left(\sum_{j=1}^{p} \beta_j x_{lj} \right) \right) \right] \right)}{\partial \beta_j}$$

$$= \sum_{i=1}^{n} \left[\sum_{j=1}^{p} x_{ij} - \frac{\sum_{l\epsilon R(t_j)} \exp\left(\sum_{j=1}^{p} \beta_j x_{lj} \right) \sum_{j=1}^{p} x_{ij}}{\sum_{l\epsilon R(t_j)} \exp\left(\sum_{j=1}^{p} \beta_j x_{lj} \right)} \right]. \tag{2.22}$$

Pendugaan β_j dapat diperoleh dengan memaksimumkan turunan pertama fungsi *log partial likelihood* yaitu dengan mencari solusi dari:

$$\frac{\partial \ln L(\beta)}{\partial \beta_j} = 0$$

$$\sum_{i=1}^{n} \left[\sum_{j=1}^{p} x_{ij} - \frac{\sum_{l\epsilon R(t_j)} \exp\left(\sum_{j=1}^{p} \beta_j x_{lj} \right) \sum_{j=1}^{p} x_{ij}}{\sum_{l\epsilon R(t_j)} \exp\left(\sum_{j=1}^{p} \beta_j x_{lj} \right)} \right] = 0. \tag{2.23}$$

Persamaan diatas dapat diselesaikan secara numerik yaitu dengan iterasi menggunakan metode *Newton-Raphson* dengan bantuan komputasi.

Turunan kedua dari $\ln L(\beta)$ terhadap β_j yaitu sebgai berikut.

$$\frac{\partial^2 \ln L(\beta)}{\partial^2 \beta_j} = \frac{\partial y}{\partial \beta_j}\left(\frac{\partial \ln L(\beta)}{\partial \beta_j}\right)$$

$$= \frac{\partial y}{\partial \beta_j}\left[\sum_{i=1}^{n}\left[\sum_{j=1}^{p} x_{ij} - \frac{\sum_{l\epsilon R(t_j)} \exp\left(\sum_{j=1}^{p} \beta_j x_{lj}\right) \sum_{j=1}^{p} x_{ij}}{\sum_{l\epsilon R(t_j)} \exp\left(\sum_{j=1}^{p} \beta_j x_{lj}\right)}\right]\right]$$

$$= \sum_{i=1}^{n}\left[\frac{\left(\sum_{l\epsilon R(t_j)}\left(\sum_{j=1}^{p} x_{ij}\right)\exp\left(\sum_{j=1}^{p} \beta_j x_{lj}\right)\right)^2}{\left(\sum_{l\epsilon R(t_j)} \exp\left(\sum_{j=1}^{p} \beta_j x_{lj}\right)\right)^2} - \frac{\sum_{l\epsilon R(t_j)} \exp\left(\sum_{j=1}^{p} \beta_j x_{lj}\right)\left(\sum_{j=1}^{p} x_{ij}\right)^2}{\sum_{l\epsilon R(t_j)} \exp\left(\sum_{j=1}^{p} \beta_j x_{lj}\right)}\right]$$

$$= -\sum_{i=1}^{n}\left[\frac{\sum_{l\epsilon R(t_j)} \exp\left(\sum_{j=1}^{p} \beta_j x_{lj}\right)\left(\sum_{j=1}^{p} x_{ij}\right)^2}{\sum_{l\epsilon R(t_j)} \exp\left(\sum_{j=1}^{p} \beta_j x_{lj}\right)} - \frac{\left(\sum_{l\epsilon R(t_j)}\left(\sum_{j=1}^{p} x_{ij}\right)\exp\left(\sum_{j=1}^{p} \beta_j x_{lj}\right)\right)^2}{\left(\sum_{l\epsilon R(t_j)} \exp\left(\sum_{j=1}^{p} \beta_j x_{lj}\right)\right)^2}\right]. \tag{2.24}$$

Negatif turunan kedua dari log likelihood yaitu sebagai berikut.

$$-\frac{\partial^2 \ln L(\beta)}{\partial^2 \beta_j}$$

$$= -\left[-\sum_{i=1}^{n}\left[\frac{\sum_{l\epsilon R(t_j)} \exp\left(\sum_{j=1}^{p} \beta_j x_{lj}\right)\left(\sum_{j=1}^{p} x_{ij}\right)^2}{\sum_{l\epsilon R(t_j)} \exp\left(\sum_{j=1}^{p} \beta_j x_{lj}\right)} - \frac{\left(\sum_{l\epsilon R(t_j)}\left(\sum_{j=1}^{p} x_{ij}\right)\exp\left(\sum_{j=1}^{p} \beta_j x_{lj}\right)\right)^2}{\left(\sum_{l\epsilon R(t_j)} \exp\left(\sum_{j=1}^{p} \beta_j x_{lj}\right)\right)^2}\right]\right]. \tag{2.25}$$

2. Prosedur *Newton Raphson*

Model untuk data *survival* tersensor biasanya menggunakan prosedur Newton Raphson dalam memaksimalkan fungsi partial likelihood. Misalkan $L(\beta)$ merupakan fungsi partial likelihood p dimensional vektor

$\boldsymbol{\beta} = (\beta_1, \beta_2, \dots, \beta_p)^t$. Misalkan $\mathbf{U}(\boldsymbol{\beta})$ merupakan vektor berukuran p dari turunan partial pertama $L(\beta)$.

$$\mathbf{U}(\boldsymbol{\beta}) = [U_1(\boldsymbol{\beta}), \dots, U_\mathrm{p}(\boldsymbol{\beta})]^\mathbf{t} \quad (2.26)$$

dengan memisalkan $U_j(\boldsymbol{\beta}) = \frac{\partial\, ln\, L(\beta)}{\partial \beta_j}, j = 1,2, \dots, p$.

Misalkan $\mathbf{I}(\boldsymbol{\beta})$ merupakan matrik Hessian berukuran $p \times p$ dari turunan partial likelihood kedua $ln\, L(\beta)$ yaitu

$$\mathbf{I}(\boldsymbol{\beta}) = \left(I_{ij}(\boldsymbol{\beta})\right), i, j = 1, \dots, p \quad (2.27)$$

dengan memisalkan $I_{ij}(\boldsymbol{\beta}) = \frac{\partial^2\, ln\, L(\beta)}{\partial \beta_i \beta_j}$

$$\mathbf{I}(\boldsymbol{\beta}) = \begin{bmatrix} \frac{\partial^2\, ln\, L(\beta)}{(\partial \beta_1)^2} & \frac{\partial^2\, ln\, L(\beta)}{\partial \beta_1 \beta_2} & \cdots & \frac{\partial^2\, ln\, L(\beta)}{\partial \beta_1 \beta_p} \\ \frac{\partial^2\, ln\, L(\beta)}{\partial \beta_2 \beta_1} & \frac{\partial^2\, ln\, L(\beta)}{\partial \beta_2 \beta_2} & \cdots & \frac{\partial^2\, ln\, L(\beta)}{\partial \beta_2 \beta_p} \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2\, ln\, L(\beta)}{\partial \beta_p \beta_1} & \frac{\partial^2\, ln\, L(\beta)}{\partial \beta_p \beta_p} & \cdots & \frac{\partial^2\, ln\, L(\beta)}{(\partial \beta_p)^2} \end{bmatrix}.$$

Algoritma metode Newton Raphon yaitu sebagai berikut.

$$\widehat{\boldsymbol{\beta}}_{\mathrm{c}+1} = \widehat{\boldsymbol{\beta}}_\mathrm{c} - \mathbf{I}(\widehat{\boldsymbol{\beta}}_\mathrm{c})^{-1}\mathbf{U}(\widehat{\boldsymbol{\beta}}_\mathrm{c}) \quad (2.28)$$

dengan memisalkan, $c = 0,1,2, \dots$ dan $\mathbf{I}^{-1}(\widehat{\boldsymbol{\beta}}_\mathrm{c})$ merupakan invers dari $\mathbf{I}(\widehat{\boldsymbol{\beta}}_\mathrm{c})$.

Langkah interasi dengan metode *Newton Raphson* sebagai berikut.

1. Menentukan nilai awal, $\widehat{\boldsymbol{\beta}}_0 = \mathbf{0}.$
2. $\widehat{\boldsymbol{\beta}}_1 = \widehat{\boldsymbol{\beta}}_0 - \mathbf{I}\left(\widehat{\boldsymbol{\beta}}_0\right)^{-1}\mathbf{U}\left(\widehat{\boldsymbol{\beta}}_0\right).$
3. Iterasi dilakukan sampai memperoleh nilai yang konvergen, $\widehat{\boldsymbol{\beta}}_{c+1} \cong \widehat{\boldsymbol{\beta}}_c.$

Varians dari $\hat{\beta}_j$ dapat didefinisikan (Hosmer, Lemeshow, & May, 2008: 72) sebagai berikut

$$Var\left(\widehat{\boldsymbol{\beta}}\right) = \mathbf{I}\left(\widehat{\boldsymbol{\beta}}\right)^{-1}. \tag{2.29}$$

Standar deviasi dari $\hat{\beta}_j$ merupakan akar kuadrat dari varians $\hat{\beta}_j$ (Hosmer, Lemeshow, & May, 2008: 72) sebagai berikut

$$SE\left(\widehat{\boldsymbol{\beta}}\right) = \sqrt{Var\left(\widehat{\boldsymbol{\beta}}\right)} = \sqrt{\mathbf{I}\left(\widehat{\boldsymbol{\beta}}\right)^{-1}}\ . \tag{2.30}$$

Standar deviasi diatas dapat digunakan untuk mencari selang kepercayaan $\hat{\beta}_j$yaitu $(1-\alpha)100\%$ selang kepercayaan untuk $\hat{\beta}_j$ ((Hosmer, Lemeshow, & May, 2008: 72) sebagai berikut

$$\hat{\beta}_j \pm z_{1-\frac{\alpha}{2}} SE\left(\widehat{\boldsymbol{\beta}}\right). \tag{2.31}$$

3. Pengujian Parameter

Melalui model Cox dapat dilihat hubungan antara variabel bebas (variabel independen) terhadap variabel terikat (variabel dependen) yaitu waktu *survival* melalui fungsi *hazard*nya, seperti yang ditunjukan pada persamaan (2.14) yaitu

$$h(t,x) = h_0(t)exp\left(\beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p\right).$$

Menurut David W. Hosmer dan Standley Lemeshow (2008: 89), terdapat tiga cara untuk menguji signifikansi parameter yaitu dengan uji *partial likelihood ratio*, uji *Wald*, dan uji *score*. Pengujian signifikasi parameter bertujuan untuk memeriksa apakah variabel bebas memiliki pengaruh nyata dalam model.

a. Uji *partial likelihood* rasio

Untuk menguji hipotesis bahwa satu atau beberapa parameter regresi β_j adalah nol dapat menggunakan uji *partial likelihood* rasio dinotasikan dengan G. Statistik uji ini mengikuti distribusi *chi-square* dengan derajat bebas p . Berikut langkah-langkah uji *partial likelihood* rasio:

1. Hipotesis:
   $H_0: \beta_j = 0$
   $H_1: \beta_j \neq 0$
2. Taraf signifikansi: α
3. Statistik uji:

$$G = -2\left[\ln L(0) - \ln L\left(\hat{\beta}_j\right)\right] \quad (2.32)$$

   Dengan memisalkan,

   $\ln L(0)$ adalah *log partial likelihood* dari model tanpa variabel bebas (model nol).

   $\ln L\left(\hat{\beta}_j\right)$ adalah *log partial likelihood* dari model yang terdiri dari p variabel bebas.

4. Daerah penolakan:

H_0 ditolak jika $G \geq \mathcal{X}^2_{(\alpha:\mathrm{db=p})}$ atau p-value $\leq \alpha$

p: banyaknya variabel bebas.

5. Kesimpulan:

Jika H_0 ditolak maka $\beta_i \neq 0$, mengindikasikan bahwa variabel bebas berpengaruh terhadap waktu *survival* (variabel dependen).

b. Uji *Wald*

Uji *Wald* digunakan untuk menguji pengaruh parameter secara terpisah, dinotasikan dengan z. Statistik uji ini mengikuti distribusi *chi-square* dengan derajat bebas p . Berikut langkah-langkah uji *Wald*:

1. Hipotesis:

   $H_0: \beta_j = 0$

   $H_1: \beta_j \neq 0$
2. Taraf signifikansi: α
3. Statistik uji:

$$z^2 = \left(\frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}\right)^2 \tag{2.33}$$

4. Daerah penolakan:

H_0 ditolak jika $z^2 \geq \mathcal{X}^2_{(\alpha:\mathrm{db=p})}$ atau p-value $\leq \alpha$

p: banyaknya variabel bebas.

5. Kesimpulan:

Jika H_0 ditolak maka $\beta_j \neq 0$, mengindikasikan bahwa variabel bebas berpengaruh terhadap waktu *survival* (variabel dependen).

c. Uji Score

Uji yang lain untuk menguji signifikasi parameter yaitu uji Score. Statistik uji ini adalah rasio dari turunan *log partial likelihood* pada persamaan (2.22), dengan akar kuadrat dari persamaan (2.25) semuanya dievaluasi terhadap $\beta_j = 0$. Statistik uji ini mengikuti distribusi *chi-square* dengan derajat bebas p . Berikut langkah-langkah uji Score:

1. Hipotesis:
   $H_0: \beta_j = 0$
   $H_1: \beta_j \neq 0$
2. Taraf signifikansi: α
3. Statistik uji:

$$z^* = \left(\left. \frac{\partial L\beta_j / \partial\beta_j}{\sqrt{\mathbf{I}(\beta_j)}} \right|_{\beta_j = 0} \right)^2 \qquad (2.34)$$

4. Daerah penolakan:

   H_0 ditolak jika $z^* \geq \mathcal{X}^2_{(\alpha:\mathrm{db=p})}$ atau p-value $\leq \alpha$

   p: banyaknya variabel bebas.

5. Kesimpulan:

   Jika H_0 ditolak maka $\beta_j \neq 0$, mengindikasikan bahwa variabel bebas berpengaruh terhadap waktu *survival* (variabel dependen).

4. Pemilihan Model Cox Terbaik

Pemilihan model terbaik diawali dengan pemilihan variabel yang masuk atau keluar dari model. Menurut David Collett (2003: 61), pemilihan variabel yang masuk atau keluar dari model dapat dilakukan dengan tiga cara yaitu seleksi *forward*, eliminasi *backward* dan prosedur *stepwise*. Prosedure seleksi *stepwise* merupakan kombinasi dari dua proses yaitu seleksi *forward* dan seleksi *backward*. Seleksi *backward* atau seleksi mundur dengan memasukkan semua variabel ke dalam model kemudian mengeluarkannya satu persatu jika variabel memiliki peningkatan nilai $-2\ln L(\beta)$ terbesar. Jika sudah tidak ada peningkatan nilai $-2\ln L(\beta)$ secara signifikan dari pengurangan variabel maka langkah *backward* dihentikan. Dalam skripsi ini pemilihan model terbaik dilakukan menggunakan seleksi *forward*. Seleksi *forward* atau seleksi maju yaitu dengan menambahkan variabel satu demi satu dalam setiap langkahnya. Menururt David W. Hosmer dan Stanley Lemeshow (2008: 416) taraf signifikansi yang digunakan dalam seleksi *forward* disarankan antara 20% - 25% untuk memungkinkan lebih banyak variabel yang masuk dalam model. Pada masing-masing tahapan, kita akan memutuskan variabel mana yang merupakan prediktor terbaik untuk dimasukkan ke dalam model. Berikut seleksi *forward*:

Langkah 0: Misalkan ada sebanyak p variabel bebas X_i, dengan $i = 1,2,3,\dots,p$. Hitung $G^{(0)}(i) = -2\left[\ln L^{(0)}(0) - \ln L^{(0)}(i)\right]$ dengan $\ln L^{(0)}(0)$ adalah *log partial likelihood* dari model nol (model tanpa variabel bebas) dan $\ln L^{(0)}(i)$ adalah *log partial likelihood* dari model dengan variabel bebas X_i, dengan *p-value* untuk uji signifikansinya yaitu $p^{(0)}(i) = \mathrm{P}\left[\chi^2(1) \geq G^{(0)}(i)\right]$. Variabel bebas yang pertama kali masuk dalam model adalah variabel bebas yang paling signifikan berpengaruh terhadap waktu *survival* dinotasikan dengan X_{e_1} yaitu variabel yang memiliki $p^{(0)}(e_1) = \min_i p^{(0)}(i)$ dengan $p^{(0)}(e_1) < \alpha$, α merupakan taraf signifikansi yang dipilih, maka proses be lanjut pada langkah 1.

Langkah 1: Langkah ini dimulai dengan variabel X_{e_1} dalam model. Hitung $G^{(1)}(i) = -2\left[\ln L^{(1)}(i) - \ln L\left(X_{e_1}\right)\right]$ dengan $i = 1,2,3,\dots,p$ dan $\neq e_1$. $\ln L\left(X_{e_1}\right)$ adalah *log partial likelihood* dengan dua varibel bebas dalam model, $\ln L^{(1)}(i)$ adalah *log partial likelihood* dengan variabel bebas yang terpilih ada langkah 0. *P-value* untuk uji signifikansi dari penambahan X_i pada model yang terdiri dari X_{e_1} yaitu $p^{(1)}(i) = \mathrm{P}\left[\chi^2(1) \geq G^{(1)}(i)\right]$. Varibel bebas yang dipilih untuk masuk dalam langkah 2 yaitu X_{e_2} dengan variabel tersebut memiliki $p^{(1)}(e_2) = \min_{i \neq e_1} p^{(1)}(i)$. Jika variabel yang terpilih yaitu X_{e_2} signifikan dengan $p^{(1)}(e_2) < \alpha$, maka berlanjut pada langkah 2.

Langkah 2: Langkah ini dimulai dengan variabel X_{e_1} dan X_{e_2} dalam model. Hitung $G^{(2)}(i) = -2\left[\ln L\left(X_{e_1}, X_{e_2}\right) - \ln L^{(2)}(i)\right]$ dengan $i = 1,2,3,\dots,p$ dan $\neq e_1, e_2$, dengan $\ln L\left(X_{e_1}, X_{e_2}\right)$adalah *log partial likelihood* dengan tiga varibel bebas dalam model, $\ln L^{(2)}(i)$ adalah *log partial likelihood* dengan variabel bebas yang terpilih pada langkah 1. P-value untuk uji signifikansi dari penambahan variabel baru pada model yang terdiri dari X_{e_1} dan X_{e_2} yaitu $p^{(2)}(i) = P[\chi^2(1) \geq G^{(2)}(i)]$. Variabel bebas yang dipilih untuk masuk dalam langkah 3 yaitu X_{e_3} dengan variabel tersebut memiliki $p^{(2)}(e_3) = \min_{i \neq e_1 e_2} p^{(2)}(i)$. Jika variabel yang terpilih yaitu X_{e_3} signifikan dengan $p^{(2)}(e_3) < \alpha$, maka berlanjut pada langkah 3.

Langkah 3: Pada langkah 3 sama dengan langkah 2 dalam proses eliminasi menentukan apakah semua variabel dimasukkan ke dalam model di langkah-langkah sebelumnya masih signifikan. Proses seleksi dilanjutkan seperti langkah sebelumnya sampai langkah terakhir yaitu langkah S.

Langkah S: Pada langkah ini, satu atau dua hal berikut terjadi yaitu: 1) semua variabel sudah masuk dalam model dan tidak ada yang keluar, 2) setiap variabel bebas tidak masuk dalam model yaitu mempunyai $p^{(s)}(i) < \alpha$. Pada langkah ini, tidak ada variabel bebas yang terpilih untuk masuk dan tidak ada variabel bebas yang keluar dari model.

Setelah diperoleh variabel yang masuk dalam model dengan beberapa langkah diatas, kemudian dilanjutkan dengan pemeriksaan apakah terdapat interaksi antar variabel tersebut dengan melakukan uji likelihood rasio dengan membandingkan model Cox tanpa interaksi dengan model Cox dengan penambahan variabel interaksi. Langkah-langkah pemilihan variabel interaksi yang masuk dalam model untuk mendapatkan model Cox terbaik dapat dilakukan dengan seleksi *forward,* seleksi *backward* maupun prosedur *stepwise* dengan langkah-langkah sama seperti yang dijelaskan di atas.

J. Residual Model *Cox Proportional Hazard*

Menurut David Collett (2003: 113) terdapat beberapa jenis residual yaitu *residual Cox-Snell, residual Martingale, residual Deviance, residual Schoenfeld* dan *residual Score*. *Residual Schoenfeld* akan dibahas dalam skripsi ini. *Residual Schoenfeld* untuk individu ke-i pada variabel bebas ke-j yaitu sebagai berikut.

$$\hat{r}_{ij} = c_i \left(x_{ij} - \frac{\sum_{l \epsilon R(t_l)} x_{ij} \exp\left(\hat{\beta}' x_{jl}\right)}{\sum_{l \epsilon R(t_l)} \exp\left(\hat{\beta}' x_{jl}\right)} \right) \tag{2.35}$$

dengan memisalkan, c_i indikator penyensoran, $c_i = 1$ untuk tidak tersensor, $c_i = 0$ untuk yang lainnya, x_{ij} adalah nilai dari variabel bebas ke-j ,$j = 1,2, \dots, p$ untuk individu ke-i dalam pengamatan. Vektor dari p *residual Schoenfeld* untuk individu ke-i yaitu sebagai berikut

$$\hat{\boldsymbol{r}}_{\boldsymbol{i}}' = \left(\hat{r}_{i1}, \hat{r}_{i2}, \dots, \hat{r}_{ip}\right). \tag{2.36}$$

Vektor dari scaled *residual Schoenfeld* yaitu sebagai berikut

$$\hat{\boldsymbol{r}}_{\boldsymbol{i}}^{*} = r \text{ var } \left(\hat{\beta}\right) \hat{r}_i \tag{2.37}$$

Dengan memisalkan, r adalah banyaknya kejadian dari n individu dan var $\left(\hat{\beta}\right)$ adalah ragam dari parameter β.

K. Pengujian Asumsi *Proportional Hazard*

Terdapat 3 cara untuk mengecek asumsi *Proportional Hazard yaitu*: dengan pendekatan grafik menggunakan plot *log minus – log survival*, dengan menggunakan *residual Schoenfeld* dan dengan menambahkan variabel dependen waktu (Collett, 2003: 179) . Pengujian asumsi *proportional hazard* yang digunakan dalam skripsi ini yaitu dengan pendekatan grafik dengan plot *log-minus-log survival* dan dengan menggunakan *residual Schoenfeld* .

Menurut David Collett (2003: 182) ada 3 pilihan untuk mengatasi *cox nonproportional hazard* yaitu mengeluarkan variabel bebas yang tidak memenuhi asumsi dari model, menggunakan model *Cox Stratifikasi* dan dengan perluasan model C*cox*.

1. Pendekatan Grafik *Log-Minus- Log Survival*

Pendekatan grafik yang digunakan yaitu dengan plot *log-minus-log survival* atau plot log kumulatif *hazard*. Menurut model regresi Cox, fungsi *hazard* untuk kejadian setiap waktu t untuk individu ke$-i$ dapat dituliskan seperti pada persamaan (2.14) yaitu sebagai berikut.

$$h_i(t,x) = h_0(t)exp\left(\beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_j x_{ji}\right)$$

$$= h_0(t)exp\left(\sum_{j=1}^{p} \beta_j x_{ji}\right)$$

$$h_i(\boldsymbol{x}_i) = h_0(t)exp(\boldsymbol{\beta x_i}). \tag{2.38}$$

Dengan memisalkan $\boldsymbol{x}_i$ merupakan vektor dari nilai variabel bebas untuk individu tersebut, $\boldsymbol{\beta}$ merupakan vektor dari parameter, dan $h_0(t)$ merupakan fungsi *hazard* dasar. Apabila kedua sisi diintegralkan dari nol hingga t, maka diperoleh sebagai berikut.

$$\int_0^t h_i(\boldsymbol{x}_i)\,dx = \exp(\boldsymbol{\beta x_i}) \int_0^t h_0(t)\,dx \qquad (2.39)$$

dengan menggunakan persamaan (2.12), sehingga diperoleh

$$H_i(t) = \exp(\boldsymbol{\beta x_i})\, H_0(t) \qquad (2.40)$$

dengan memisalkan $H_i(t)$ dan $H_0(t)$ merupakan fungsi kumulatif *hazard*. Selanjutnya dilakukan logaritma pada persamaan (2.40) untuk kedua sisi:

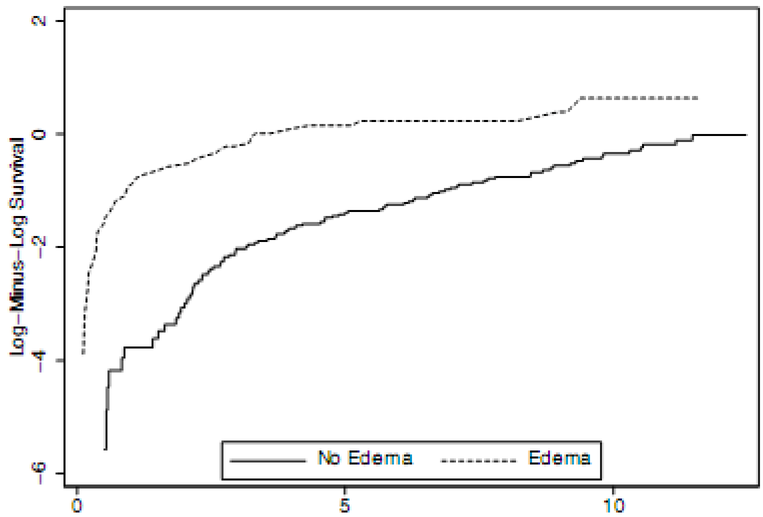
$$\log H_i(t) = \log \exp(\boldsymbol{\beta x_i}) + \log H_0(t). \qquad (2.41)$$

Dengan menggunakan persamaan (2.13), didapatkan:

$$\log\left[-\log\left(S_i(t)\right)\right] = \exp(\boldsymbol{\beta x_i}) + \log\left[-\log\left(S_0(t)\right)\right]. \qquad (2.42)$$

Dari persamaan (2.42) menunjukkan bahwa fungsi *log minus log survival* tidak bergantung terhadap waktu. Ini berarti bahwa fungsi *log minus log survival* pada model *Cox proportional hazard* pada persamaan (2.14) valid jika diplotkan terhadap berlawanan waktu *survival* maka kurva yang terbentuk akan parallel. Dalam menggunakan plot *log minus log survival* ini, data *survival* dikelompokkan sesuai dengan tingkat dari satu atau lebih faktor. Jika variabel kontinu maka nilainya perlu dikelompokkan menjadi variabel kategori. Plot *log minus log* *survival* adalah sebuah plot dari logaritma estimasi fungsi kumulatif *hazard* terhadap waktu *survival*, akan menghasilkan kurva paralel jika laju *proportional hazard* diseluruh kelompok yang berbeda. Kelemahan menggunakan pendekatan grafik/plot adalah bersifat subyektif; paralel atau tidaknya tergantung cara pandang peneliti (Collett, 2003: 179).

Sebagai contoh misalnya pada kasus pengaruh edema (peningkatan volume cairan pada kaki) pada penderita primary biliary cirrhosis(PBC) yaitu penyakit kerusakan saluran-saluran kecil empedu di hati, yang menyebabkan empedu menumpuk di hati. Berikut adalah gambar plot *log- minus-log survival* untuk variabel edema (Vittinghoff, Glidden, Shiboski, & McCulloch, 2005: 235).



Gambar 2. 1 Plot *log-minus-log survival* pada variabel edema

Pada Gambar 2.1 tersebut memperlihatkan bahwa plot log-minus- log *survival* pada pasien dengan edema dan tidak mengalami edema mendekati paralel, sehingga mengindikasikan bahwa asumsi *proportional hazard* pada variabel edema terpenuhi (Vittinghoff, Glidden, Shiboski, & McCulloch, 2005: 235).

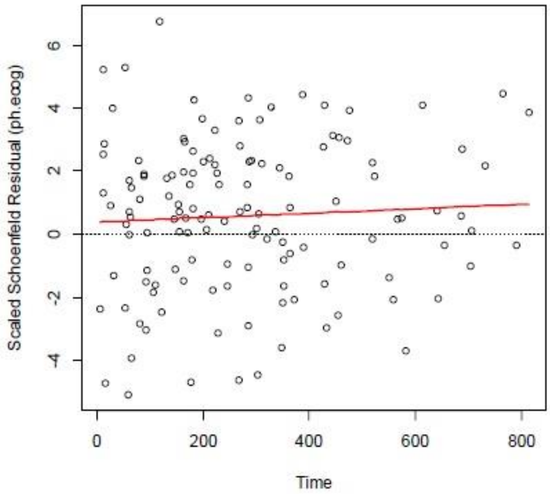
2. Menggunakan *Residual Schoenfeld*

Cox Proportional Hazard dikatakan *proportional* jika rasio dari *hazard* independen terhadap waktu. Jika terdapat variabel bebas yang dependen waktu maka asumsi *proportional hazard* tidak terpenuhi. Residual Schoenfeld dapat digunakan untuk mengecek asumsi *proportional hazard*. Therneau & Grambsch (2000: 137) menunjukkan bahwa nilai harapan dari *scaled residual Schoenfeld* ke-i, untuk variabel bebas ke-j $\left(X_j\right)$ dalam model yaitu sebagai berikut:

$$\mathrm{E}\left(\mathrm{r}^*_{\mathrm{pij}}\right) \approx \beta_j\left(t_j\right) - \hat{\beta}_j \tag{2.43}$$

dengan memisalkan $\beta_j(t)$ adalah nilai dari parameter pada saat waktu kejadian ke-i. $\hat{\beta}_j$ merupakan nilai estimasi dari $\beta_j$pada model regresi cox. Plot dari nilai $\mathrm{r}^*_{\mathrm{pij}} + \hat{\beta}_j$ berlawanan terhadap waktu kejadian memberikan informasi tentang bentuk dari koefisien dependen waktu dari X_j yaitu $\beta_j(t)$. Apabila plot tersebut horisontal maka mengindikasikan bahwa koefisien dari X_j konstan, dan asumsi *proportional hazard* terpenuhi.

Sebagai contoh misalnya pada kasus penyakit kanker paru-paru yang diperoleh dari *North Central Cancer Treatment Group* (NCCTG) yang terdiri dari 228 data pasien dari sebuah studi variabel prognostik tentang perkembangan kanker paru-paru. Berikut adalah gambar plot *Residual Schoenfeld* pada variabel nilai *ecog* (Therneau & Grambsch, 2000: 137).



Gambar 2. 2 plot *Residual Schoenfeld* pada variabel nilai ecog

Pada gambar 2.2 tersebut memperlihatkan bahwa *Residual Schoenfeld* pada variabel nilai *ecog* memiliki kemiringan mendekati nol atau mendekati horisontal, sehingga mengindikasikan bahwa asumsi *proportional hazard* untuk variabel nilai *ecog*.

L. Interpretasi Model Regresi Cox

Persamaan regresi cox $h(t, \boldsymbol{X_j}) = h_0 t \exp(\beta \boldsymbol{X_j})$ dapat diinterpretasi sebagai berikut.

Untuk himpunan kovariat x_0 dan x_1 dari dua individu maka diperoleh

$$\left[\frac{h(t, \boldsymbol{X_j})}{h(t, \boldsymbol{X_0})}\right] = \frac{h_0 t \exp(\beta X_1)}{h_0 t \exp(\beta X_0)} = \frac{\exp(\beta X_1)}{\exp(\beta X_0)} = e^{(X_1 - X_0)\beta}, \forall \mathrm{t} > 0. \qquad (2.44)$$

Persamaan (2.44) menunjukkan besarnya rasio relatif dari individu dengan faktor risiko X_1 dibandingkan dengan faktor risiko X_1 dari individu lain (Lee,2003: 134).

Persamaan (2.44) dapat dituliskan

$$\log\left[\frac{h(t,X_1)}{h(t,X_0)}\right] = e^{(X_1-X_0)\beta}$$

$$\log\left[\frac{h(t,X_1)}{h(t,X_0)}\right] = \log\left(e^{(X_1-X_0)\beta}\right)$$

$$\log\left[\frac{h(t,X_1)}{h(t,X_0)}\right] = (X_1 - X_0)\beta. \tag{2.45}$$

Untuk setiap kenaikan $\boldsymbol{X}_j$ sedangkan nilai kovariat yang lainnya tetap dapat diinterpretasikan

$$\log\left[\frac{h\left(t,\boldsymbol{X_{j+1}}\right)}{h(t,\boldsymbol{X_0})}\right] = \beta_j. \tag{2.46}$$

Persamaan (2.46) dapat disimpulkan bahwa setiap naiknya nilai β_j akan memperbesar nilai *log hazard rasio.* Sehingga dapt dituliskan

$$\log\left[\frac{h\left(t,\boldsymbol{X_{j+1}}\right)}{h(t,\boldsymbol{X_0})}\right] = \beta_j$$

$$\left[\frac{h\left(t,\boldsymbol{X_{j+1}}\right)}{h(t,\boldsymbol{X_0})}\right] = e^{\beta_j}\ , \quad \forall \mathrm{t} > 0. \tag{2.47}$$

Dengan demikian nilai $\exp\left(\beta_j\right)$ merupakan hazard rasio yang dapat dihubungkan dengan kenaikan nilai x_j.

Karena $h\left(t,\boldsymbol{X_j}\right) \approx \mathrm{P}(t < T < t + \Delta t | T \geq t, \boldsymbol{X})$ maka persamaan (2.47) dapat dituliskan

$$\frac{\mathrm{P}\left(t < T < t + \Delta t \middle| T \geq t, \boldsymbol{X}_{j+1}\right)}{\mathrm{P}\left(t < T < t + \Delta t \middle| T \geq t, \boldsymbol{X}_{j}\right)} = e^{\beta_j}\,, \quad \forall \mathrm{t} > 0. \tag{2.48}$$

Jadi $\exp(\beta_j)$ dapat juga diinterpretasikan sebagai rasio 2 probabilitas bersyarat dari gagalnya individu yang diketahui tersebut masih hidup pada saat t. Sehingga persamaan (2.47) ekuivalen dengan

$$\left[\frac{h(t, \boldsymbol{X_{j+1}}) - h(t, \boldsymbol{X_j})}{h(t, \boldsymbol{X_j})}\right] = e^{\beta_j} - 1\ , \forall \mathrm{t} > 0\,. \tag{2.49}$$

Sehingga $e^{\beta_j} - 1$ dapat diinterpretasikan sebagai persentase perubahan nilai *hazard* baik naik atau turun dari setiap naiknya nilai x_j, dengan menganggap kovariat yang lain tetap.

Terdapat 3 macam ketentuan tentang bertambahnya atau berkurangnya nilai *hazard* sebagai berikut.

1. $\beta_j > 0$ maka setiap naiknya nilai x_j akan memperbesar nilai hazard atau semakin besar risiko seorang individu untuk meninggal atau gagal.
2. $\beta_j < 0$ maka setiap naiknya nilai x_j akan memperkecil nilai hazard atau semakin kecil risiko seorang individu untuk meninggal atau gagal.
3. $\beta_j = 0$ maka besar risiko seorang individu untuk hidup sama dengan besarnya risiko seorang individu untuk meninggal atau gagal.

BAB III
HASIL DAN PEMBAHASAN

Pada Bab III akan dibahas tentang prosedur prosedur pemodelan *Cox proportional hazard* pada kasus kejadian bersama dan penerapan pemodelan *Cox proportional hazard* pada kasus kejadian bersama. Terlebih dahulu akan dijelaskan mengenai kejadian bersama.

A. Kejadian Bersama

Dalam *analisis survival* terkadang ditemukan adanya kejadian bersama atau yang sering disebut *ties*. *Ties* adalah keadaan yang terdapat dua individu atau lebih yang mengalami kejadian pada waktu yang bersamaan. Jika suatu data terdapat *ties*, maka akan menimbulkan permasalahan dalam membentuk *partial likelihoodnya* yaitu saat menentukkan anggota dari himpunan risikonya.

Sebagai contoh untuk menggambarkan kejadian bersama dalam kejadian, akan digunakan tabel 3.1, dengan memisalkan i adalah individu ke-i dan t_i adalah waktu kejadian untuk individu ke-i.

i	1	2	3	4
t_i	5	5	8	14

Tabel 3.1 Data *Survival* dengan terdapat *ties*

Misalkan $t_1 < t_2 < t_3 < t_4$ adalah waktu yang teramati yang telah diurutkan. Pada waktu $t = 5$, terdapat dua objek yang mengalami kejadian dan tidak diketahui objek mana yang mengalami kejadian terlebih dahulu. Kejadian bersama tersebut dapat menimbulkan permasalahan pada estimasi parameter yang berhubungan dengan penentuan anggota dari himpunan risiko. Banyak metode dalam mengestimasi parameter pada kasus kejadian bersama, salah satunya dengan pendekatan metode *Breslow*. Metode *Breslow* mengasumsikan bahwa ukuran dari himpunan risiko untuk kejadian bersama adalah sama, selengkapnya akan dijelaskan pada prosedur estimasi parameter.

B. Prosedur Pemodelan *Cox proportional hazard* Pada Kejadian Bersama

Berdasarkan kajian pustaka pada bab II mengenai analisis *survival* dan *Cox proportional hazard* secara umum, maka akan dijelaskan proses pemodelan *Cox proportional hazard* pada kasus kejadian bersama. Langkah-langkah dalam pemodelan *Cox proportional hazard* pada kasus kejadian bersama sebagai berikut.

1. Identifikasi Data

   Proses identifikasi data dalam analisis survival sangat penting, dimana waktu survival serta variabel bebas dan terikat yang akan digunakan dalam penelitian ini ditentukan diawal proses analisis.

2. Estimasi Parameter Model *Cox*

   Pada estimasi digunakan pendekatan metode *besrlow*. Pendekatan ini banyak digunakan karena fungsi *partial likelihoodnya* sederhana daripada metode lain (Breslow, 1974: 54). Dalam setiap kasus kejadian bersama tidak mungkin untuk menentukan urutan kejadian, metode *Breslow* mengasumsikan bahwa ukuran dari himpunan risiko adalah sama. Terdapat dua kasus yang memiliki waktu yang sama yaitu 5 yang dapat dilihat pada tabel 3.1. Urutan kejadian antara individu 1 dan individu 2 tidak dapat dibedakan dan ketiga kejadian tersebut tidak saling mempengaruhi atau saling bebas (independen). Berdasarkan persamaan (2.19) dapat disusun bentuk *partial likelihood* untuk individu 1 sebagai berikut.

$$\mathrm{P}(\mathrm{A}_1|\mathrm{B}) = \frac{\exp(\beta x_1)}{\exp(\beta x_1) + \exp(\beta x_2)}\,. \tag{3.1}$$

Menurut Breslow (1974: 54), maka himpunan risiko untuk individu 2 sama dengan himpunan risiko untuk individu 1, sehingga bentuk *partial likelihood* untuk individu 2 sebagai berikut.

$$\mathrm{P}(\mathrm{A}_2|\mathrm{B}) = \frac{\exp(\beta x_2)}{\exp(\beta x_1) + \exp(\beta x_2)}\,. \tag{3.2}$$

Dari persamaan (3.1) dan (3.2) memberikan fungsi hazard dasar sebagai berikut.

$$\mathrm{P}(\mathrm{A}|\mathrm{B}) = \frac{\exp(\beta x_1)}{\exp(\beta x_1) + \exp(\beta x_2)} \times \frac{\exp(\beta x_2)}{\exp(\beta x_1) + \exp(\beta x_2)}$$

$$= \frac{\exp(\beta x_1) \times \exp(\beta x_2)}{(\exp(\beta x_1) + \exp(\beta x_2))^2}$$

$$= \frac{\exp(\beta x_1 + \beta x_2)}{(\sum_{\mathrm{i}=1}^{2} \exp(\beta x_i))^2}$$

$$= \frac{\exp\big(\beta(x_1 + x_2)\big)}{(\sum_{\mathrm{i}=1}^{2} \exp(\beta x_i))^2}\,. \tag{3.4}$$

Dari persamaan (3.4) diperoleh bentuk umum dari fungsi hazard dasar sebagai berikut.

$$\mathrm{P(A|B)} = \frac{\exp\left(\sum_{j=1}^{p} \beta_j S_k\right)}{\left(\sum_{l \epsilon R(t_j)} \exp\left(\sum_{j=1}^{p} \beta_j X_{lj}\right)\right)^{d_i}} . \tag{3.5}$$

Dengan S_k adalah jumlah kovarian $\boldsymbol{x}$ pada kasus *ties* dan d_i adalah banyaknya kasus ties pada waktu t_i. Dengan mengambil fungsi hazard dasar (3.5), memberikan fungsi *partial likelihood* sebagai berikut.

$$L(\beta)_{breslow} = \prod_{i=1}^{r} \frac{\exp\left(\sum_{j=1}^{p} \beta_j S_k\right)}{\left(\sum_{l \epsilon R(t_j)} \exp\left(\sum_{j=1}^{p} \beta_j X_{lj}\right)\right)^{d_i}} \tag{3.6}$$

3. Pemilihan Model Terbaik

Menurut David Collett (2003: 61), pemilihan variabel yang masuk atau keluar dari model dapat dilakukan dengan tiga cara yaitu seleksi *forward*, eleminasi *backward* dan prosedur *stepwise*. Pada penelitian ini menggunakan seleksi *forward*, sehingga masing-masing tahapan akan diputuskan variabel mana yang merupakan prediktor terbaik untuk dimasukkan ke dalam model.

4. Pengujian Parameter

Terdapat tiga cara dalam pengujian parameter untuk menguji signifikansi parameter yaitu dengan uji *partial likelihood ratio,* uji wald dan uji score. Dalam penelitian ini untuk mengetahui variabel-variabel yang berpengaruh signifikan dalam pembentukan model *Cox proportional hazard*, maka dilakukan pengujian setiap variabel dengan uji *wald*.

5. Penyusunan Model *Cox* pada kasus bersama

Setelah dilakukan seleksi model terbaik maka akan diperoleh model terbaik, yang akan digunakan sebagai model terakhir C*ox proportional hazard.*

6. Pengujian Asumsi *Proportional Hazard*

Pengujian Asumsi *Proportional Hazard* sangatlah penting karena untuk mengetahui rasio fungsi hazard dari dua individu konstan dari waktu ke waktu atau ekuivalen dengan pernyataan bahwa fungsi hazard suatu individu terhadap fungsi hazard individu yang lain adalah proporsional. Pengujian ini dengan menggunakan pendekatan *plot minus log survival* dan *residual Schoenfeld.* Jika semua variabel bebas memenuhi asumsi *proportional hazard* maka model akhir regresi *Cox* disebut model *Cox proportional hazard*. Apabila sebaliknya ada varibel bebas yang tidak memenuhi asumsi ini, maka model regresi *Cox* terebut disebut model *Cox nonproportional hazard.* Dalam model *Cox nonproportional hazard* terdapat perbedaan fungsi *hazard* dari satu individu terhadap yang lain. (Guo, 2009: 82). Menurut David Collett (2003: 182) ada 3 pilihan untuk mengatasi *Cox nonproportional hazard* yaitu mengeluarkan variabel bebas yang tidak memenuhi asumsi proportional hazard dari model, menggunakan model *Cox stratifikasi* dan dengan perluasan model *Cox*.

7. Interpretasi Model *Cox* pada kasus bersama

Proses interpretasi merupakan proses terakhir dimana menjelaskan serta menyimpulkan hasil dari model terakhir yang digunakan dalam *Cox proportional hazard.*

C. Penerapan Pemodelan *Cox proportional hazard* Pada Kejadian Bersama

Rata-rata kecelakaan di Amerika Serikat yang mencapai 50.000 kasus kecelakaan dalam 1 tahun. Data kecelakaan lalu lintas di Amerika tahun 2012 yang diambil dari website *National Highway Traffic Safety Administration* (nhtsa.gov) terdapat 44.862 kasus kecelakaan. Peraturan lalu lintas yang berlaku ketat di Amerika Serikat belum mengurangi kasus kecelakaan lalu lintas. Oleh sebab itu, akan dilakukan analisis untuk mengetahui faktor-faktor yang menyebabkan kecelakaan lalu lintas di Amerika Serikat. Pada penelitian ini digunakan 398 data orang yang mengalami kecelakaan pertama kali pada tanggal 1 Januari 2010. Data ditampilkan secara lengkap pada lampiran 1. Waktu pengamatan dimulai dari 1 Januari 2010 sampai 31 Desember 2012. Penentuan waktu awal dan akhir pengamatan mengakibatkan terdapatnya waktu kecelakaan yang sama atau dapat dikatakan kejadian bersama. Berdasarkan uraian diatas, kasus kecelakaan lalu lintas di Amerika dapat dianalisis menggunakan model *Cox.* Kejadian bersama dalam kasus kecelakaan lalu lintas di Amerika Serikat akan dianalisis dengan pendekatan metode Breslow.

1. Identifikasi Data

Variabel yang dimasukkan dalam analisis yaitu variabel umur, jenis kelamin, kepemilikan SIM, penggunaan sabuk pengaman, pengaruh alkohol saat mengemudi.. Identifikasi variabel data kecelakaan lalu lintas sebagain berikut.

a. Variabel terikat

Waktu hidup pengemudi yang diukur dari bulan kecelakaan pertama kali (1 Januari 2010) sampai bulan terjadinya kecelakaan kedua. Ada dua kondisi yang dapat terjadi pada pengemudi pada kejadian kecelakaan kedua, pengemudi meninggal dunia atau pengemudi masih hidup dan mengalami penyensoran setelah kecelakaan kedua. (1 untuk tersensor dan 2 untuk pengemudi yang meninggal).

b. Variabel bebas

a. Umur (X_1)

Umur pengemudi pada saat mengalami kecelakaan.

b. Jenis kelamin (X_2)

Dalam mengemudikan kendaraan terdapat perbedaan gaya atau sikap saat mengemudi antara seorang perempuan dengan laki-laki, dimana perempuan cenderung berhati-hati dari laki-laki maupun sebaliknya (Rinto, 2014: 73). Oleh karena itu, variabel jenis kelamin pengemudi dibedakan menjadi dua yaitu 1 = laki-laki dan 2 = perempuan.

c. Kepemilikan SIM (X_3)

Dalam mengemudi kepemilikan Surat Ijin Mengemudi (SIM) sangatlah diperlukan, dimana SIM tersebut sebagai penanda dimana seorang telah mengetahui dan menguasai cara berkendara yang baik. Akan tetapi banyak seseorang mendapatkan SIM secara illegal atau dengan cara instan sehingga banyak kecelakaan dikarenakan pengemudi tidak pandai dalam berkendara yang baik. Kepemilikan SIM pengemudi, bentuknya berupa variabel kategorik yaitu 0 = memiliki SIM dan 1 = tidak memiliki SIM.

d. Pemakaian sabuk pengaman (X_4)

Sabuk pengaman saat sangatlah penting dimana faktor ini berpengaruh sangat besar dalam mengurangi benturan saat terjadinya kecelakaan. Pengemudi sangat terbantu dengan sabuk ini dimana saat terjadinya benturan yang sangat besar di kendaraan secara otomatis sabuk tersebut mengencang dan mengikat secara erat pengemudi pada tempat duduknya. Pemakaian sabuk pengaman saat mengemudi, bentuknya berupa variabel kategorik yaitu 0 = memakai sabuk pengaman dan 1 = tidak memakai sabuk pengaman.

e. Pengaruh alkohol saat mengemudi (X_5)

Sering kali kecelakaan disebabkan oleh pengemudi mengonsumsi minuman alkohol secara berlebihan sehingga mempengaruhinya saat berkendara seperti timbulnya halusinasi atau kurangnya konsentrasi saat berkendara. Pengaruh alkohol saat mengemudi, bentuknya berupa variabel kategorik yaitu 0 = tidak mengkonsumsi alkohol dan 1 = mengkonsumsi alkohol.

2. Estimasi Parameter Model *Cox*

Parameter β_j pada model *Cox* merupakan parameter yang belum diketahui dan akan diduga menggunakan metode maximum *partial likelihood* estimasi *breslow*. Dengan bantuan *software* R versi 3.0.3 diperoleh estimasi parameter dengan metode *breslow* untuk setiap variabel pada data kecelakaan lalu lintas (output selengkapnya pada lampiran 2) sebagai berikut.

Tabel 3.2 Estimasi Parameter Model *Cox* dengan metode *Breslow*

Variabel	Koef	$exp(\beta_j)$	SE	$p > \|z\|$	95% CI dari hazard rasio	
					Batas bawah	Batas atas
Umur	0,012404	1,012481	0,004705	0,00838	1,0032	1,0219
Jenis kelamin	0,221632	1,248112	0,173730	0,20205	0,8879	1,7544
SIM	-0,536257	0,584934	0,166715	0,00130	0,4219	0,811
Sabuk Pengaman	1,131677	3.100851	0,169729	2,6e-11	2,2233	4,325
Alkohol	0,188945	1,207974	0,158201	0,23235	0,8859	1,6471

Diasumsikan semua variabel berpengaruh terhadap model, maka semua variabel dimasukkan dalam persamaan umum model cox, sehingga diperoleh estimasi model *Cox proportional hazard* dengan metode *breslow* sebagai berikut.

$$h(t,\boldsymbol{X}) = h_0(t)exp(0{,}012404\, X_1 + 0{,}221632\, X_2 - 0{,}536257\, X_3 + 1{,}131677\, X_4 + 0{,}188945\, X_5)\,. \quad (3.7)$$

Guna mengetahui apakah model diatas sudah tepat, maka dilakukan uji *partial likelihood ratio* sebagai berikut.

- Hipotesis:

  $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = 0$ (variabel X_1, X_2, X_3, X_4, X_5 tidak berpengaruh dalam model)

  $H_1: \exists \beta_i \neq 0, i = 1{,}2{,}3{,}4{,}5$ (variabel X_1, X_2, X_3, X_4, X_5 berpengaruh dalam model)

- Taraf signifikansi: $\alpha = 0{,}05$
- Statistik uji:

  $$G = -2[lnL_R - lnL_F]$$

- Daerah penolakan: H_0 ditolak jika $G \geq \mathcal{X}^2_{(0,05:5)}$ atau *p-value* < 0,05
- Perhitungan

  Dari hasil *output software* R yang ditampilkan dalam lampiran 2, diperoleh nilai *log likelihood* untuk model *Cox* tanpa variabel bebas (model *null*) yaitu $lnL_R = -980{,}4182$ dan nilai *log likelihood* untuk model *Cox* pada persamaan (3.2) yaitu $lnL_F = -947{,}4668$. sehingga diperoleh perhitungan sebagai berikut.

  $$G = -2[lnL_R - lnL_F]$$

$$= -2[-980{,}4182 - (-947{,}4668)]$$

$$= 65{,}9028\,.$$

Karena $G = 65{,}9028 \geq \mathcal{X}^2_{(0,05:5)} = 11{,}070$ dan *p-value*=7,280843e-13 < 0,05, sehingga H_0 ditolak dan dapat disimpulkan bahwa variabel X_1, X_2, X_3, X_4, X_5 berpengaruh dalam model, mengindikasikan bahwa model pada persamaan (3.2) lebih baik dari pada model tanpa variabel bebas (model *null*).

3. Pemilihan Model *Cox* terbaik

Menurut David Collett (2003, 61), pemilihan variabel yang masuk atau keluar dari model dapat dilakukkan dengan tiga cara yaitu seleksi *forward*, eleminasi *backward* dan prosedur *stepwise*. Pada penelitian ini menggunakan seksi *forward*, sehingga masing-masing tahapan akan diputuskan variabel mana yang merupakan prediktor terbaik untuk dimasukkan ke dalam model.

Pemilihan model terbaik pada tabel di bawah ini diperoleh model dengan *p-value* terkecil dari setiap langkah. Proses penambahan variabel bebas berhenti pada langkah ke empat karena semua variabel sudah masuk ke dalam model. Berikut. langkah-langkah pemilihan model terbaik dengan seleksi *forward*.

Tabel 3.3 Prosedur Seleksi *Forward* dalam Pemilihan Model Terbaik

	Model	LL (β)	$G(i)$	*p-value*
Langkah 0	0	-980,4182		
	X_1	-979,0736	2,689213	0,1010297
	X_2	-980,4059	0,02464925	0,8752442

	X_3	-977,2083	6,419729	0,01128594
	$\boldsymbol{X_4}$	**-957,9863**	**44,86372**	**2,112355e-11**
	X_5	-978,5448	3,746807	0,05290849
Langkah 1	$X_4 + X_1$	-954,1608	7,651116	0,00567372
	$X_4 + X_2$	-956,8910	2,190641	0,1388515
	$\boldsymbol{X_4 + X_3}$	**-951,9841**	**12,00438**	**0,000530758**
	$X_4 + X_5$	-957,9672	0,03825415	0,8449337
Langkah 2	$\boldsymbol{X_4 + X_3 + X_1}$	**-948,9244**	**6,119519**	**0,01336971**
	$X_4 + X_3 + X_2$	-951,0706	1,827004	0,1764818
	$X_4 + X_3 + X_5$	-951,7158	0,5365999	0,4638452
Langkah 3	$X_4 + X_3 + X_1 + X_2$	-948,1732	1,502356	0,2203092
	$X_4 + X_3 + X_1 + X_5$	-948,2519	1,344945	0,2461638
Langkah 4	$X_4 + X_3 + X_1 + X_2 + X_5$	-947,4668	1,412858	0,2345826

Dengan memisalkan,
X_1 : umur, X_4 : sabuk pengaman,
X_2: jenis kelamin, X_5 : alkohol,
X_3 : SIM.

Langkah 0: Pemilihan model terbaik berikut. menggunakan taraf signifikansi yaitu $\alpha = 0{,}05$. Pada langkah 0 yang tersaji dalam table 3.3 yang selengkapnya pada lampiran 3 memperlihatkan bahwa dari perhitungan nilai $G^{(0)}(i)$ dan *p-value*, model dengan penambahan variabel X_i dengan $i = 1{,}2{,}3{,}4{,}5$ memiliki nilai $p^{(0)}(i) < 0{,}05$. Dari perhitungan tersebut diperoleh $G^{(0)}(i)$ terbesar yaitu $G^{(0)}(4) = 44.86372$ dan mempunyai *p-value* terkecil yaitu $p^{(0)}(4) =$ 2,112355e-11 sehingga X_4

paling signifikan berpengaruh dengan waktu *survival*. Variabel X_4 terpilih untuk dimasukkan ke dalam langkah berikut.nya, yaitu langkah 1.

Berikut analisisnya,

- Hipotesis:

  $H_0 : \beta_4 = 0$ (variabel sabuk pengaman tidak berpengaruh dalam model)

  $H_1 : \exists \beta_4 \neq 0$ (variabel sabuk pengaman berpengaruh dalam model)

- Taraf signifikansi: $\alpha = 0{,}05$

- Statistik uji:

$$G^{(0)}(i) = -2\left[lnL^{(0)}(0) - lnL^{(0)}(i)\right]$$

- Daerah penolakan: H_0 ditolak jika $G^{(0)}(i) \geq \mathcal{X}^2_{(0{,}05:1)}$ atau *p-value* < 0,05

- Perhitungan:

  Dari hasil *output software* R pada tabel 3.3 yang selengkapnya pada lampiran 3 diperoleh $lnL^{(0)}(0) = -980{,}4182$ dan $lnL^{(0)}(4) = -957.9863$ dengan *p-value* dari uji *log partial likelihood* adalah 2,112355e-11. Sehingga nilai dari uji *log partial likelihood ratio* dengan perhitungan sebagai berikut.

$$\begin{aligned} G^{(0)}(4) &= -2\left[lnL^{(0)}(0) - lnL^{(0)}(4)\right] \\ &= -2[-980{,}4182 - (-957{,}9863)] \\ &= 44{,}86372. \end{aligned}$$

  Karena $G = 44{,}86372 \geq \mathcal{X}^2_{(0{,}05:1)} = 3{,}841$ dan *p-value* = 2,112355e-11 $\leq \alpha = 0{,}05$, sehingga H_0 ditolak dan dapat disimpulkan bahwa sabuk pengaman ada hubungan dengan waktu *survival*. Variabel X_4 paling berpengaruh dengan waktu *survival* sehingga variabel tersebut dimasukkan dalam model.

Langkah 1: Pada langkah ini dimulai dengan memasukkan variabel X_4 dalam model. Variabel X_4 dimodelkan dengan menambah satu persatu variabel X_1, X_2, X_3, X_5 kemudian dilakukan uji *partial likelihood ratio* dengan membandingkan model terdiri 2 varibel tersebut dengan model yang terdiri dari satu variabel yaitu X_4, hasil perhitungan disajikan pada tabel 3.3 yang selengkapnya pada lampiran 3 pada langkah 1. Variabel yang terpilih untuk masuk pada langkah 2 yaitu variabel yang memiliki *p-value* terkecil yaitu model dengan penambahan variabel X_3 dengan $p - value = 0.000530758$. Variabel X_3 dan X_4 terpilih dan dimasukkan dalam langkah 2.

Berikut analisisnya,

- Hipotesis:

  $H_0: \beta_3 = \beta_4 = 0$ (variabel X_3 dan X_4 tidak berpengaruh dalam model)

  $H_1: \exists\beta_i \neq 0, i = 3,4$ (variabel X_3 dan X_4 berpengaruh dalam model)

- Taraf signifikansi: $\alpha = 0,05$
- Statistik uji:

$$G^{(1)}(i) = -2\left[lnL(X_{ei}) - lnL^{(1)}(0)\right]$$

- Daerah penolakan: H_0 ditolak jika $G^{(1)}(i) \geq \mathcal{X}^2_{(0,05:1)}$ atau *p-value* < 0,05
- Perhitungan:

Dari hasil *output software* R pada tabel 3.3 yang selengkapnya pada lampiran 3 diperoleh $lnL^{(1)}(3) = -951{,}9841$ dan $lnL(X_4) = -957{,}9863$ dengan *p-value* dari uji *log partial likelihood* adalah 0,000530758. Sehingga nilai dari uji *log partial likelihood ratio* dengan perhitungan sebagai berikut.

$$G^{(1)}(3) = -2\left[lnL(X_4) - lnL^{(1)}(3)\right]$$

$$= -2[-957{,}9863 - (-951{,}9841)]$$

$$= 12{,}00438.$$

Karena $G^{(1)}(3) = 12{,}00438 \geq \mathcal{X}^2_{(0{,}05:1)} = 3{,}841$ dan *p-value* = 0,000530758$\leq \alpha = 0{,}05$, sehingga H_0 ditolak dan dapat disimpulkan bahwa variabel X_3 dan X_4 berpengaruh dalam model sehingga variabel X_3 dan X_4 disertakan pada langkah selanjutnya yaitu langkah 2.

Langkah 2: Pada langkah ini dimulai dengan memasukkan variabel X_3 dan X_4 dalam model. Variabel X_3 dan X_4 dimodelkan dengan menambah satu persatu variabel X_1, X_2, X_5 kemudian dilakukan uji *partial likelihood ratio* dengan membandingkan model terdiri 3 varibel tersebut dengan model yang terdiri dari dua variabel yaitu X_3 dan X_4, hasil perhitungan disajikan pada tabel diatas pada langkah 2. Variabel yang terpilih untuk masuk pada langkah 3 yaitu variabel yang memiliki *p-value* terkecil yaitu model dengan penambahan variabel X_1 dengan $p - value = 0{,}01336971$. Variabel X_3 dan X_4 terpilih dan dimasukkan dalam langkah 3.

Berikut analisisnya,

- Hipotesis:

  $H_0: \beta_1 = \beta_3 = \beta_4 = 0$ (variabel X_1, X_3 dan X_4 tidak berpengaruh dalam model)

  $H_1: \exists\beta_i \neq 0, i = 1,3,4$ (variabel X_1, X_3 dan X_4 berpengaruh dalam model)

- Taraf signifikansi: $\alpha = 0{,}05$
- Statistik uji:

$$G^{(2)}(i) = -2\left[lnL(X_{ei}) - lnL^{(2)}(0)\right]$$

- Daerah penolakan: H_0 ditolak jika $G^{(1)}(i) \geq \mathcal{X}^2_{(0,05:1)}$ atau *p-value* < 0,05
- Perhitungan:

  Dari hasil *output software* R pada tabel 3.3 yang selengkapnya pada lampiran 3 diperoleh $lnL^{(2)}(1) = -948{,}9244$ dan $lnL(X_3, X_4) = -951{,}9841$ dengan *p-value* dari uji *log partial likelihood* adalah $0{,}01336971$. Sehingga nilai dari uji *log partial likelihood ratio* dengan perhitungan sebagai berikut.

$$G^{(2)}(1) = -2\left[lnL(X_3, X_4) - lnL^{(2)}(1)\right]$$

$$= -2[-951{,}9841 - (-948{,}9244)]$$

$$= 6{,}119519.$$

  Karena $G^{(2)}(1) = 6{,}119519 \geq \mathcal{X}^2_{(0,05:1)} = 3{,}841$ dan *p-value* = $0{,}01336971 \leq \alpha = 0{,}05$, sehingga H_0 ditolak dan dapat disimpulkan bahwa variabel X_1, X_3 dan X_4 berpengaruh dalam model sehingga variabel X_1, X_3 dan X_4 disertakan pada langkah selanjutnya yaitu langkah 3.

Langkah 3: Pada langkah ini dimulai dengan memasukkan variabel X_1, X_3 dan X_4 dalam model. Variabel X_1, X_3 dan X_4 dimodelkan dengan menambah satu persatu variabel X_2, X_5 kemudian dilakukan uji *partial likelihood ratio* dengan membandingkan model terdiri 4 varibel tersebut dengan model yang terdiri dari tiga variabel yaitu X_1, X_3 dan X_4, hasil perhitungan disajikan pada tabel diatas pada langkah 3. Variabel yang terpilih untuk masuk pada langkah 4 yaitu variabel yang memiliki *p-value* terkecil yaitu model dengan penambahan variabel X_2 dengan $p-value = 0{,}2203092$. Variabel X_1, X_3 dan X_4 terpilih dan dimasukkan dalam langkah 4.

Berikut analisisnya,

- Hipotesis:

  $H_0: \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ (variabel X_1, X_2, X_3 dan X_4 tidak berpengaruh dalam model)

  $H_1: \exists \beta_i \neq 0, i = 1,2,3,4$ (variabel X_1, X_2, X_3 dan X_4 berpengaruh dalam model)

- Taraf signifikansi: $\alpha = 0{,}05$

- Statistik uji:

$$G^{(3)}(i) = -2\left[lnL(X_{ei}) - lnL^{(3)}(0)\right]$$

- Daerah penolakan: H_0 ditolak jika $G^{(1)}(i) \geq \mathcal{X}^2_{(0,05:1)}$ atau *p-value* < 0,05

- Perhitungan:

Dari hasil *output software* R pada tabel 3.3 diperoleh $lnL^{(3)}(2) = -948{,}1732$ dan $lnL(X_1, X_3, X_4) = -948{,}9244$ dengan *p-value* dari uji *log partial likelihood* adalah $0{,}2203092$. Sehingga nilai dari uji *log partial likelihood ratio* dengan perhitungan sebagai berikut.

$$G^{(3)}(2) = -2\left[lnL(X_1, X_3, X_4) - lnL^{(3)}(2)\right]$$

$$= -2[-948{,}9244 - (-948{,}1732)]$$

$$= 1{,}502356.$$

Karena $G^{(3)}(2) = 1{,}502356 \geq \mathcal{X}^2_{(0{,}05:1)} = 3{,}841$ dan *p-value* $= 0{,}2203092 \geq \alpha = 0{,}05$, sehingga H_0 diterima dan dapat disimpulkan bahwa variabel X_2 tidak berpengaruh dalam model yang terdiri dari variabel X_1, X_3 dan X_4

Langkah 4: Karena pada langkah 3, variabel X_2 ditolak maka secara otomatis variabel X_5 juga tidak berpengaruh terhadap model yang terdiri dari variabel X_1, X_3 dan X_4. Semua variabel sudah dimasukkan dalam model maka seleksi *forward* telah selesai

Berdasarkan hasil dari seleksi *forward* didapatkan tiga variabel terpilih yang masuk dalam model terbaik *Cox proportional hazard* yaitu umur, kepemilikan SIM, dan pemakaian sabuk pengaman. Tabel dibawah menampilkan hasil estimasi parameter model terbaik *Cox proportional hazard* berdasarkan hasil seleksi *forward* yang selengakapnya pada lampiran 3.

Tabel 3.4 Estimasi Parameter model *Cox* Terbaik dengan Seleksi *Forward*

Variabel	Koef	$exp(\beta_j)$	SE	$p > \|z\|$	95% CI dari hazard rasio	
					Batas bawah	Batas atas
Umur	0,011687	1,011755	0,004609	0,01122	1,0027	1,0209
SIM	-0,517437	0,596046	0,164976	0,00171	0,4314	0,8236
Sabuk Pengaman	1,134048	0,321728	0,162828	3,29e-12	2,2590	4,2767

Model *Cox proportional hazard* berdasarkan hasil seleksi *forward* yaitu sebagai berikut.

$$h(t, \boldsymbol{X}) = h_0(t) exp(0{,}011687\ X_1 - 0{,}517437\ X_3 + 1{,}134048\, X_4). \qquad (3.8)$$

Langkah terakhir dalam pemilihan model terbaik yaitu seleksi interaksi dengan seleksi *forward*. Tabel dibawah ini menunjukkan daftar pasangan variabel yang berinteraksi sebagi calon variabel yang akan masuk dalam model, dimulai dengan menganggap model *Cox proportional hazard* pada persamaan diatas sebagai model *null* (model tanpa interaksi). Langkah-langkah pemilihan model interaksi terbaik dengan seleksi *forward* sama dengan langkah-langkah dalam pemilihan model terbaik pada persamaan diatas. Berikut. langkah-langkah dalam pemilihan model terbaik dengan interaksi:

Langkah 0: Dalam pemilihan model terbaik berikut. menggunakan taraf signifikansi yaitu $\alpha = 0{,}05$. Langkah 0 yang terjadi dalam tabel dibawah memperlihatkan dari perhitungan nilai $G^{(0)}(i)$ dan *p-value*, semua model dengan penambahan satu variabel interaksi *p-value*nya lebih besar dari

$\alpha = 0{,}05$, yang berarti bahwa tidak ada variabel interaksi yang signifikan berpengaruh pada model. Sehingga tidak ada variabel interaksi yang terpilih untuk masuk pada langkah berikut.nya.

Langkah 1: Karena setiap variabel bebas tidak ada yang masuk dalam model yaitu mempunyai *p-value* kurang dari $\alpha = 0{,}05$. Pada langkah ini, tidak ada variabel bebas yang terpilih untuk masuk, maka seleksi *forward* telah selesai.

Berdasakan hasil seleksi *forward* dalam pemilihan variabel interaksi untuk mendapatkan model *Cox proportional hazard* terbaik yang disajikan dalam tabel dibawah, dapat disimpulkan tidak ada variabel interaksi yang signifikan pada taraf 5%. Sehingga model terbaik berdasarkan hasil seleksi *forward* yaitu model *Cox proportional hazard* pada persamaan (3.8).

Tabel 3.5 Prosedur Seleksi *Forward* Pemilihan Interaksi Model Terbaik

	Model	LL (β)	$G(i)$	*p-value*
Langkah 0	0	-948,9244		
	$(X_1 \times X_3)$	-948,7877	0,2733147	0,6011169
	$(X_1 \times X_4)$	-948,3273	1,194121	0,2744999
	$(X_3 \times X_4)$	-948,8633	0,1222255	0,7266332

Dengan memisalkan,

X_1 : umur,

X_3: SIM,

X_4 : sabuk pengaman.

4. Pengujian Parameter

Dalam pengujian parameter terdapat tiga cara untuk menguji signifikansi parameter yaitu dengan uji *partial likelihood ratio*, uji wald dan uji score. Dalam penelitian ini untuk mengetahui variabel-variabel yang berpengaruh signifikan dalam pembentukan model *Cox proportional hazard*, maka dilakukan pengujian setiap variabel dengan uji wald. Uji wald dilakukan pada tiga yang telah masuk dalam persamaan diatas yaitu variabel umur, SIM dan sabuk pengaman. Hasil pengujian parameter secara parsial menggunakan uji wald dengan bantuan *software* R pada tabel 3.6 yang selengkapnya pada lampiran 4 yaitu sebagai berikut.

- Hipotesis:

  $H_0: \beta_j = 0$ (variabel X_j tidak berpengaruh terhadap waktu *survival*)

  $H_1: \exists \beta_j \neq 0, j = 1,3,4$ (variabel berpengaruh terhadap waktu *survival*)
- Taraf signifikansi: $\alpha = 0{,}05$
- Statistik uji: Uji Wald
- Daerah Penolakan: H_0 ditolak jika $z^2 \geq \chi^2_{(0,05,1)}$ atau *p-value* < 0,05.

Tabel 3.6 Hasil Pengujian Parameter Secara Partial dengan Uji Wald

Variabel	Koef	SE	z^2_{hitung}	$\mathcal{X}^2_{(0,05:1}$	Uji wald *p-value*	keputusan
Umur	0,007793	0,004669	2,785875	3,841	0,09509	H_0 diterima
SIM	-0,4025	0,1632	6,082631	3,841	0,01363	H_0 ditolak
Sabuk Pengaman	1,0170	0,1606	40,1006	3,841	2,423e-10	H_0 ditolak

Berdasarkan tabel 3.6 dapat disimpulkan sebagai berikut.

1) Variabel umur tidak berpengaruh terhadap waktu *survival*, hal tersebut dapat dilihat dari *p-value* dari uji wald pada tabel diatas

yaitu 0,09509 > $\alpha = 0{,}05$ atau $z^2 = 2{,}785875 < \chi^2_{(0{,}05{,}1)} = 3{,}841$ maka H_0 diterima,

2) Variabel SIM berpengaruh terhadap waktu *survival*, hal tersebut dapat dilihat dari *p-value* dari uji wald pada tabel diatas yaitu 0,01363 $\leq \alpha = 0{,}05$ atau $z^2 = 6{,}082631 \geq \chi^2_{(0{,}05{,}1)} = 3{,}841$ maka H_0 ditolak.

3) Variabel sabuk pengaman berpengaruh terhadap waktu *survival*, hal tersebut dapat dilihat dari *p-value* dari uji wald pada tabel diatas yaitu 2,423e-10$\leq \alpha = 0{,}05$ atau $z^2 = 40{,}1006 \geq \chi^2_{(0{,}05{,}1)} = 3{,}841$ maka H_0 ditolak.

Dari hasil uji Wald diatas menunjukkan bahwa satu variabel tidak berpengaruh pada model pada tingkat signifikansi 5% yaitu variabel umur, sehingga variabel tersebut dikeluarkan dari model persamaan. Terdapat dua variabel yang signifikan berpengaruh dalam pembentukan model *Cox proportional hazard* yaitu SIM dan sabuk pengaman. Sehingga diperoleh estimasi parameter untuk model dengan dua variabel yaitu SIM dan sabuk pengaman sebagai berikut.

Tabel 3.7 Estimasi Parameter dengan Dua Variabel yang Signifikan

Variabel	Koef	$exp(\beta_j)$	SE	$p > \lvert z \rvert$	95% CI dari hazard rasio	
					Batas bawah	Batas atas
SIM	-0,5495	0,5772	0,1641	0,000812	0,4185	0,7962
Sabuk Pengaman	1,0848	2,9589	0,1617	1,95e-11	2,1553	4,0622

Berdasarkan tabel 3.7 diperoleh model *Cox proportional hazard* sebagai berikut.

$$h(t, \boldsymbol{X}) = h_0(t)exp(-0{,}5495\, X_3 + 1{,}0848\, X_4\,). \qquad (3.9)$$

5. Penyusunan Model *Cox proportional hazard*

Selanjutnya dilakukan uji *partial likelihood* antara model pada persamaan 3.9 dengan model persamaan 3.4 untuk mengetahui model mana yang dipilih sebagai model akhir *Cox proportional hazard*. Langkah-langkah uji *partial likelihood* sebagai berikut.

- Hipotesis:

  $H_0: \beta_1 = 0$ (*Model reduce*)

  $H_1: \beta_1 \neq 0$ (*Model full*)

- Taraf Signifikansi: $\alpha = 0{,}05$
- Statistik uji:

  $$G = -2[lnL_R - lnL_F]$$

  Dengan:

  lnL_R merupakan *log partial likelihood ratio model reduce* (model pada persamaan (3.9))

  lnL_F merupakan *log partial likelihood ratio model full* (model pada persamaan (3.8))

- Daerah penolakan: H_0 ditolak jika $G \geq \mathcal{X}^2_{(0{,}05:5)}$ atau *p-value* < 0,05
- Perhitungan:

Dari hasil *output software* R yang selengkapnya pada Lampiran 4 diperoleh *log partial likelihood* dari *model full* yaitu $lnL_F = -948,9244$ dan *log partial likelihood* dari *model reduce* yaitu $lnL_R = -951,9841$.

$$G = -2[lnL_R - lnL_F]$$
$$= -2[-951,9841 — 948,9244)$$
$$= 6,1194$$

Nilai *p-value* dari uji *log partial likelihood* tersebut yaitu $P\left(\chi^2_{(0,05,1)} \geq (6.1194)\right) = 0,01336971.$ Karena $G = 6,1194 \geq \chi^2_{(0,05,1)} = 3,841$ dan $p - value = 0,01336971 < \alpha = 0,05$ sehingga H_0 ditolak, hal ini mengindikasikan bahwa variabel umur berkontribusi dalam model yang terdiri dari variabel SIM dan sabuk pengaman. Dengan kata lain bahwa model pada persamaan (3.8) lebih baik dari pada model pada persamaan (3.9). Sehingga model pada persamaan (3.8) dipilih sebagai model akhir *Cox proportional hazard*.

6. Pengujian Asumsi *Proportional Hazard*

Asumsi terpenting yang harus dipenuhi dalam regresi *Cox* yaitu asumsi *proportional hazard* yang berarti bahwa rasio fungsi *hazard* dari dua individu konstan dari waktu ke waktu atau ekuivalen dengan pernyataan bahwa fungsi *hazard* suatu individu terhadap fungsi *hazard* individu yang lain adalah *proportional*. Pengujian asumsi *proportional hazard* dengan beberapa pendekatan sebagai berikut.

a. Pendekatan Grafik dengan plot *log-minus-log survival*

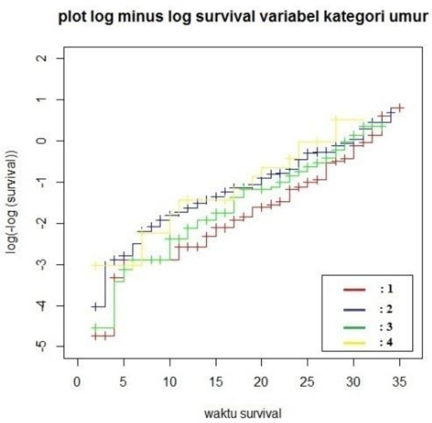
Pendekatan grafik yang digunakan yaitu dengan plot *log minus log survival*. Menurut David Collett (2003: 179) plot *log minus log survival* yang paralel menunjukkan bahwa asumsi *proportional hazard* tidak dilanggar. Dalam pendekatan grafik dengan plot *log minus log survival* terdapat kelemahan yaitu dalam menentukan paralel atau tidaknya plot tergantung dari pandangan peneliti. Berikut. pengujian asumsi *proportional hazard* dengan menggunakan plot *log minus log survival* pada masing-masing variabel bebas dengan bantuan *software* R, selengkapnya dapat dilihat pada pada Lampiran 5.

1) Pengujian asumsi *proportional hazard* dengan plot *log minus log survival* pada variabel umur

Menurut David Collett (2003: 179) jika variabel bebas yang dianalisis menggunakan plot *log minus log survival* merupakan variabel kontinu maka terlebih dahulu dilakukan pengelompokan sehingga menjadi variabel kategori. Karena umur pengemudi ini merupakan variabel kontinu dengan kisaran nilai yaitu dari 16 sampai dengan 100, maka dilakukan pengelompokan umur berdasarkan karakteristik sifat pengemudi menjadi empat kelompok umur yaitu sebagai berikut.

$$\text{Kategori umur}\begin{cases} 1, \text{umur } 16-25 \text{ tahun} \\ 2, \text{umur } 26-44 \text{ tahun} \\ 3, \text{umur } 45-64 \text{ tahun} \\ 4, \ \text{umur } 65-100 \text{ tahun} \end{cases}$$

Pengemudi yang memiliki umur 16-25 tahun cenderung mengemudi tidak stabil dikarenakan belum adanya pengalaman yang lebih, untuk pengemudi memiliki umur 26-44 tahun cenderung mengemudi sangatlah tanggungjawab karena sudah memiliki kemapanan hidup. Pengemudi yang memiliki umur 45-64 adalah jenis pengemudi yang berhati-hati dalam mengemudi, sedangkan pengemudi yang berumur 65-100 tahun merupakan jenis pengemudi yang sudah tidak dianjurkan untuk mengemudikan kendaraan (Rinto,2014: 73).



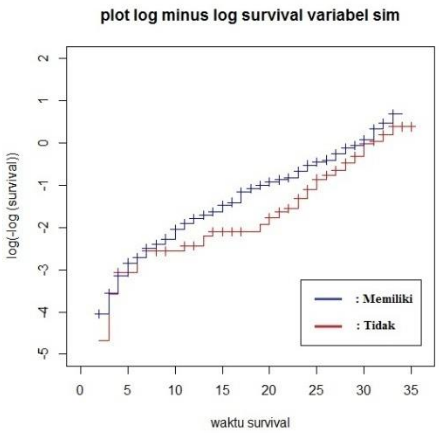
.

Gambar 3. 1 Plot *log minus log survival* untuk variabel umur

Gambar 3.1 berikut. menampilkan plot *log minus log survival* untuk variabel umur. Gambar tersebut menunjukkan bahwa plot dari kategori umur 1,2 dan 3 mendekati paralel, sedangkan plot kategori umur 4 bersilangan dengan lainnya.

Menurut David Collett (2003, 179) bentuk plot tersebut dapat dikatakan memiliki sedikit alasan untuk meragukan plot asumsi *proportional hazard s*ehingga diperlukan uji asumsi yang lainnya.

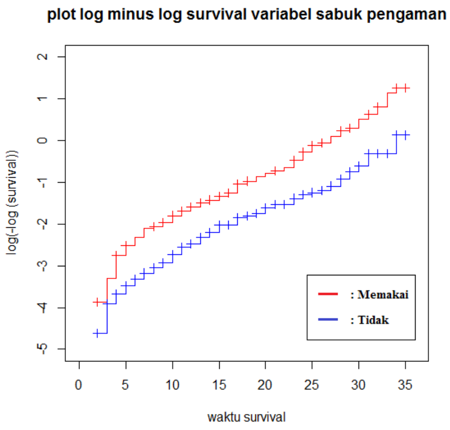
2) Pengujian asumsi *proportional hazard* dengan plot *log minus log survival* pada variabel SIM



Gambar 3. 2 Plot *log minus log survival* untuk variabel SIM

Plot *log minus log survival* untuk variabel SIM pada Gambar 3.2 menunjukkan bahwa plot untuk pengemudi yang memiliki dan yang tidak memiliki SIM berpotongan. Sehingga dapat dikatakan bahwa variabel SIM tidak memenuhi asumsi *proportional hazard.*

3) Pengujian asumsi *proportional hazard* dengan plot *log minus log survival* pada variabel sabuk pengaman



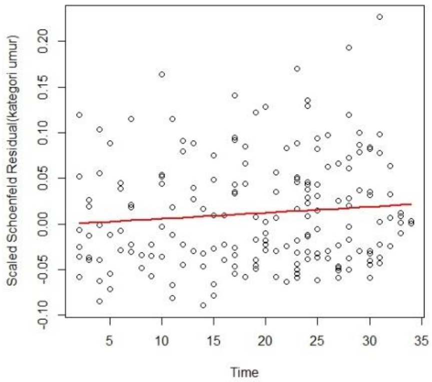
Gambar 3. 3 Plot *log minus log survival* untuk variabel Sabuk Pengaman

Plot *log minus log survival* untuk variabel Sabuk Pengaman pada Gambar 3.3 menunjukkan bahwa plot untuk pengemudi yang menggunakan sabuk pengaman dan yang tidak menggunakan sabuk pengaman tidak berpotongan dan mendekati paralel.Sehingga dapat dikatakan bahwa variabel sabuk pengaman memenuhi asumsi *proportional hazard*.

b. Menggunakan *Residual Schoenfeld*

Menurut Therneau & Grambsch (2000: 137), residual Shoenfeld dapat digunakan untuk menguji asumsi *proportional hazard* yaitu dengan *scaled residual Schoenfeld*. Plot nilai dari *scaled residual Shoenfeld* terhadap waktu *survival* memberikan informasi tentang bentuk koefisien yang dependen waktu. Kurva mendekati horisontal atau kurva memiliki kemiringan mendekati nol mengindikasikan bahwa koefisien dari X_j konstan dan asumsi *proportional hazard* terpenuhi. Berikut. *plot scaled residual Schoenfeld* terhadap waktu *survival* dari masing- masing variabel (output pada Lampiran 5):

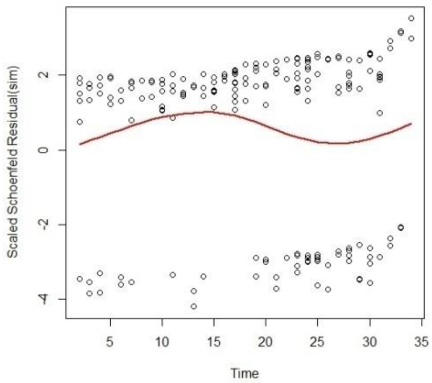
1) Variabel kategori umur



Gambar 3. 4 Plot *residual schoenfeld* untuk variabel Kategori Umur

Berdasarkan gambar 3.4 tersebut terlihat bahwa plot *scaled Schoenfeld residual* terhadap waktu *survival* dari variabel kategori umur mendekati horizontal atau memiliki kemiringan mendekati nol sehingga dapat dikatakan bahwa asumsi *proportional hazard* terpenuhi untuk variabel kategori umur.

2) Variabel SIM

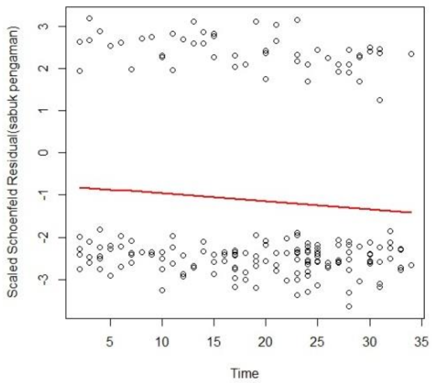


Gambar 3. 5 Plot *residual schoenfeld* untuk variabel SIM

Berdasarkan gambar 3.5 tersebut terlihat bahwa plot *scaled Schoenfeld residual* terhadap waktu *survival* dari variabel SIM tidak mendekati horizontal sehingga dapat dikatakan bahwa asumsi *proportional hazard* tidak terpenuhi untuk variabel SIM.

3) Variabel Sabuk Pengaman

Berdasarkan gambar 3.6 terlihat bahwa plot *scaled Schoenfeld residual* terhadap waktu *survival* dari variabel kategori Sabuk Pengaman mendekati horizontal atau memiliki kemiringan mendekati nol sehingga dapat dikatakan bahwa asumsi *proportional hazard* terpenuhi untuk variabel Sabuk Pengaman



Gambar 3. 6 Plot *residual schoenfeld* untuk variabel Sabuk Pengaman

Berdasarkan pengujian asumsi *proportional hazard* dengan pendekatan grafik menggunakan plot *log minus log survival* terhadap waktu *survival* dapat disimpulkan bahwa terdapat dua variabel yang secara kuat memenuhi asumsi *proportional hazard* yaitu variabel umur dan sabuk pengaman sedangkan variabel SIM tidak memenuhi asumsi *proportional hazard*. Sehingga perlu dilakukan pengujian dengan pendekatan grafik menggunakan plot *scaled Schoenfeld residual* terhadap waktu *survival*. Setelah dilakukan pengujian menggunakan plot *scaled Schoenfeld residual* dapat disimpulkan bahwa terdapat dua variabel yang secara kuat memenuhi asumsi *proportional hazard* yaitu variabel umur dan sabuk pengaman sedangkan variabel SIM tidak memenuhi asumsi *proportional hazard*. Karena variabel SIM tidak memenuhi asumsi *proportional hazard* maka variabel tersebut dikeluarkan dari model.

7. Interpretasi model *Cox proportional hazard*

Berdasarkan uji log *partial likelihood* dan pengujian asumsi *proportional hazard* disimpulkan bahwa model akhir *Cox proportional hazard* sebagai berikut.

$$h(t,\boldsymbol{X}) = h_0(t)exp(0{,}011687\ X_1 + 1{,}134048\ X_4)\ . \qquad (3.10)$$

Persamaan (3.10) menunjukan nilai $\exp(\beta_j)$ menunjukan pengaruh variabel terikat terhadap fungsi *hazard* sebagai berikut.

a. Setiap bertambahnya umur pengemudi maka semakin menambah besar risiko yang dimiliki oleh pengemudi untuk mengalami kecelakaan lalu lintas berikut.nya, yaitu sebesar $e^{0{,}011687} = 1{,}011756$, maka bertambahnya umur mengakibatkan risiko kecelakaannya sangat kecil yaitu $|(1{,}011756 - 1)100\%| = 1{,}17\%$.

b. Setiap bertambahnya pengemudi yang tidak menggunakan sabuk pengaman maka semakin menambah besar risiko yang dimiliki oleh pengemudi untuk mengalami kecelakaan lalu lintas berikut.nya, yaitu sebesar $e^{1{,}134048} = 3{,}108213$, maka besarnya risiko yang dimiliki oleh pengemudi pengemudi yang tidak menggunakan sabuk pengaman sangat besar yaitu $|(3{,}108213 - 1)100\%| = 210{,}82\%$. Dengan kata lain, pengemudi tidak menggunakan sabuk pengaman 3 kali lebih berisiko daripada pengemudi yang menggunakan sabuk pengaman.

BAB IV
PENUTUP

A. Kesimpulan

Berdasarkan hasil pembahasan mengenai *cox proportional hazard* untuk kejadian berulang pada kejadian bersama maka dapat disimpulkan sebagai berikut.

1. Prosedur pembentukan model *Cox* pada kasus kejadian bersama yakni melalui beberapa langkah yaitu sebagai berikut.
   a. Identifikasi Data.
   b. Estimasi Parameter setiap variabel dengan memaksimumkan fungsi *partial likelihood* dengan pendekatan metode *Breslow*.
   c. Pemilihan model terbaik dengan menggunakan metode *forward procedure.*
   d. Pengujian Parameter untuk mengetahui variabel bebas yang berpengaruh signifikan menggunakan Uji *Wald.*
   e. Penyusunan model *Cox* pada kasus kejadian bersama.
   f. Pengujian Asumsi *Proportional Hazard.*
      1) Dengan menggunakan pendekatan grafik plot *log-minus log survival.*
      2) Dengan menggunakan *residual Schoenfeld.*
   g. Interpretasi model *Cox* pada kasus kejadian bersama.

2. Pada contoh penerapan data, digambarkan persoalan tentang bagaimana pengaruh dari variabel umur, jenis kelamin, kepemilikan Surat Ijin Mengemudi (SIM), penggunaan sabuk pengaman, dan pengaruh alkohol terhadap waktu *survival* sampai pengemudi mengalami kecelakaan lalu lintas yang berikutnya. Dari analisis diperoleh model cox yaitu sebagai berikut

$$h(t, \boldsymbol{X}) = h_0(t) exp(0{,}011687 \ X_1 + 1{,}134048 \, X_4 \,)$$

dapat disimpulkan sebagai berikut.

c. Setiap bertambahnya umur pengemudi maka semakin menambah besar risiko yang dimiliki oleh pengemudi untuk mengalami kecelakaan lalu lintas berikutnya, yaitu sebesar $e^{0{,}011687} = 1{,}011756$, maka bertambahnya umur mengakibatkan risiko kecelakaannya sangat kecil yaitu $1{,}17\%$.

d. Setiap bertambahnya pengemudi yang tidak menggunakan sabuk pengaman maka semakin menambah besar risiko yang dimiliki oleh pengemudi untuk mengalami kecelakaan lalu lintas berikutnya, yaitu sebesar $e^{1{,}134048} = 3{,}108213$, maka besarnya risiko yang dimiliki oleh pengemudi pengemudi yang tidak menggunakan sabuk pengaman sangat besar yaitu $210{,}82\%$. Dengan kata lain, pengemudi tidak menggunakan sabuk pengaman 3 kali lebih berisiko daripada pengemudi yang menggunakan sabuk pengaman.

B. Saran

Adapun saran yang dapat diberikan untuk pengembangan dalam penelitian selanjutnya yaitu sebagai berikut.

1. Bagi penelitian lain yang akan melakukan penelitian selanjutnya mengenai cara mengatasi kasus kejadian bersama (ties) dapat dilakukan dengan pendekatan lainnya, misalnya efron.
2. Penelitian selanjutnya dapat diterapkan model *cox proportional hazard* untuk kejadian berulang pada kasus kejadian untuk data kecelakaan di Indonesia.

DAFTAR PUSTAKA

Ahmad, W. (2008). *Pengaruh Penggunaan Sabuk Pengaman (Safety Belt) Terhadap Tingkat Fatalitas Kecelakaan dan Tingkat Keparahan Kecelakaan.* Tesis. Semarang: UNDIP.

Collett, D. (2003). *Modelling Survival Data in Medical Research.* US: Chapman & Hall.

Cox, D., & Oakes, D. (1984). *Analysis of Survival Data.* London: Chapman & Hall

Fauzi Rahmawati. (2013). *Model Cox Stratifikasi Untuk Mengatasi Non Proportional Hazard Pada Analisis Survival.* Skripsi. Yogyakarta: UNY

Guo, S. (2009). *Survival Analysis.* New York: Oxford University Press.

Haddon, W. (1980). Advances in the Efidemiology of Injuries as a Basic of Public Policy. *Public Health Report* , 411-421.

Haryono, S. (1993). *Interaksi Faktor-Faktor Penyebab Kecelakaan Lalu Lintas di Jalan Tol Sekitar Jakarta.* Tesis. Jakarta: UI.

Hosmer, D. W., Lemeshow, S., & May, S. (2008). *Applied Survival Analysis: Regression Modelling of Time to Event Data.* New Jersey: John Wiley.

Klein, J. P., & Moeschberger, M. L. (2003). *Survival Analysis: Techiques for Censored and Truncated Data Second Edition.* New York: Springer-Verlag.

Kleinbaum, D. G., & Klein, M. (2005). *Survival Analysis: A Self-Learning Text Second Edition.* USA: Springer Science+Business Media, Inc.

Nisa, Shofa F. dan Budiantara, I Nyoman. 2012. Analisis *Survival* dengan Pendekatan *Multivariate Adaptive Regression Splines* pada Kasus Demam Berdarah Dengue (DBD). Jurnal Sains dan Seni ITS, Vol. 1, No. 1

Lawless, J. F. (2007). *The Statistical Analysis of Recurrent Event.* USA: Springer Science+Business Media, INc.

Lay, M. G. (1986). *Handbooks of Road Technology First Edition.* New York: Gordon and Breach Science Publishers SA.

Lee, E. T., & Wang, J. W. (2003). *Statistical Methods for Survival Data Analysis Third Edition.* New Jersey: John Wiley & Sons, Inc.

Lui, K & Pollock, D. (1991). *An Application of Proportional Hazards Model to Study the Recurrent Time Between Traffic Accidents or Infractions and Subsequent Fatal Automobile Crashes, 1986-1988*. *Journal of Safety Research* Vol. 22.

Polri, Ditlantas.(2009). *Sosialisasi UU Nomor 22 Tahun 2009 Tentang Lalu Lintas dan Angkutan Darat.* Jakarta

Rinto, R. (2014). *Tertib Berlalu Lintas.* Yogyakarta: Shafa Media.

Syahwan Udin. (2010). *Regresi Cox*. Skripsi. Yogyakarta: UNY.

Therneau, T. M., & Grambsch, P. M. (2000). *Modeling Survival Data Extending The Cox Modek.* New York: Springer_Verlag.

Vittinghoff, E., Glidden, D. V., Shiboski, S. C., & McCulloch, C. E. (2004). *Regression Methods in Biostatistics First Edition.* New York: Springer Science+Business Media.

LAMPIRAN

Lampiran 1. Data pengemudi yang mengalami kecelakaan pada tanggal 1 januari 2010 di Amerika Serikat

Keterangan:

age : Umur pengemudi saat terjadi kecelakaan.

sex : Jenis kelamin pengemudi. (1: laki-laki dan 2: perempuan)

l_status : Kepemilikan SIM. (0: memiliki dan 1: tidak memiliki)

rest_use : Penggunaan Sabuk Pengaman.
(0: menggunakan dan 1: tidak menggunakan)

dr_drink : Pengemudi mengonsumsi alkohol saat mengemudi kendaraan.
(0: tidak mengomsumsi dan 1: mengonsumsi)

durasi : Selang waktu dari 1 Januari sampai kecelakaan berikutnya.

status : Kondisi pengemudi setelah terjadi kecelakaan
(1: hidup/ tersensor dan 2: meninggal dunia)

kategori umur : Pengelompokkan umur
(1 : 16-25 tahun, 2 : 26-44 tahun, 3 : 45-64 tahun dan 4 : > 65 tahun

ID	age	sex	l_status	rest_use	dr_drink	durasi	status	kategori umur
1	50	1	1	0	0	21	2	3
2	21	1	0	0	0	14	1	1
3	21	1	0	0	0	15	2	1
4	49	1	0	1	0	4	2	3
5	48	1	0	0	0	17	2	3
6	27	1	0	0	1	18	1	2
7	20	1	1	0	0	5	1	1
8	43	1	0	1	1	17	1	2
9	55	1	1	0	1	29	2	3
10	43	1	1	1	1	23	1	2
11	55	1	0	1	0	15	1	3
12	23	1	1	0	0	31	1	1
13	65	1	0	1	0	28	2	4
14	19	1	1	1	0	35	1	1
15	30	2	0	0	0	27	1	2
16	21	1	0	1	0	11	1	1
17	56	1	0	0	0	14	1	3

ID	age	sex	l_status	rest_use	dr_drink	durasi	status	kategori umur
18	24	1	1	0	1	23	2	1
19	25	2	0	0	0	25	1	1
20	44	1	0	0	0	32	1	2
21	31	1	0	1	0	13	1	2
22	36	2	0	0	0	22	1	2
23	59	1	0	0	0	2	1	3
24	34	2	0	0	0	22	1	2
25	20	2	0	1	0	27	2	1
26	43	1	1	1	1	23	2	2
27	37	2	0	0	0	5	2	2
28	21	1	1	1	1	23	2	1
29	20	2	0	0	0	24	1	1
30	56	1	0	1	0	18	2	3
31	22	2	1	1	0	4	2	1
32	52	2	0	0	0	23	2	3
33	68	1	0	0	1	2	1	4
34	19	1	0	0	0	13	1	1
35	24	1	1	1	1	23	2	1
36	27	2	1	0	1	28	1	2
37	26	1	1	1	1	22	2	2
38	56	1	0	0	0	26	1	3
39	22	2	1	1	1	19	2	1
40	53	1	1	0	0	30	2	3
41	29	1	1	0	1	21	2	2
42	40	2	0	0	0	20	1	2
43	52	1	1	0	1	26	1	3
44	28	1	0	0	0	28	2	2
45	30	1	0	0	0	2	1	2
46	37	2	0	0	0	28	1	2
47	22	1	0	0	0	30	2	1
48	30	1	1	0	1	7	1	2
49	20	1	0	0	0	32	1	1
50	35	1	1	0	0	29	1	2
51	26	1	0	0	0	31	1	2
52	21	1	0	0	0	23	1	1
53	53	2	1	0	0	29	2	3
54	23	2	0	1	1	26	1	1
55	24	1	1	0	1	19	2	1

ID	age	sex	l_status	rest_use	dr_drink	durasi	status	kategori umur
56	19	1	1	1	0	27	2	1
57	41	1	0	0	0	25	1	2
58	22	2	0	0	0	28	1	1
59	40	1	0	1	0	7	2	2
60	31	2	0	0	0	17	1	2
61	37	1	0	0	0	20	2	2
62	30	1	0	1	0	16	2	2
63	25	2	0	1	1	8	1	1
64	56	2	1	0	0	24	1	3
65	22	2	0	0	0	21	1	1
66	25	1	0	1	1	26	2	1
67	56	1	1	0	0	8	1	3
68	38	1	0	1	0	15	2	2
69	31	1	0	0	0	22	1	2
70	42	1	0	1	1	13	2	2
71	28	1	0	0	0	8	2	2
72	51	1	1	0	1	31	1	3
73	35	1	0	1	0	17	1	2
74	59	1	1	0	1	26	2	3
75	19	1	0	0	1	26	1	1
76	48	1	0	1	0	16	1	3
77	55	1	0	1	1	12	2	3
78	71	2	0	0	0	12	1	4
79	27	1	0	1	0	9	2	2
80	28	1	0	0	0	31	1	2
81	28	1	0	1	0	7	2	2
82	48	1	1	0	0	33	1	3
83	41	1	0	0	0	22	1	2
84	68	2	0	0	0	11	2	4
85	27	1	0	0	0	13	1	2
86	22	1	1	1	1	24	1	1
87	20	1	0	0	0	23	1	1
88	22	2	0	0	1	29	1	1
89	32	1	0	1	1	2	1	2
90	43	1	0	0	1	11	1	2
91	52	1	0	0	0	17	1	3
92	26	1	0	0	0	31	2	2
93	30	1	0	0	0	11	1	2

ID	age	sex	l_status	rest_use	dr_drink	durasi	status	kategori umur
94	35	1	0	1	1	32	2	2
95	48	1	0	0	0	29	1	3
96	23	1	0	1	0	31	2	1
97	29	2	0	1	1	17	2	2
98	22	1	0	0	1	6	1	1
99	35	1	0	1	1	4	2	2
100	38	1	1	1	0	3	2	2
101	21	1	1	1	1	30	2	1
102	23	1	0	1	1	5	1	1
103	20	1	0	0	1	11	2	1
104	29	1	0	1	0	15	2	2
105	36	1	1	0	0	16	1	2
106	37	1	0	1	1	21	2	2
107	26	2	0	0	0	31	2	2
108	30	1	0	1	0	7	2	2
109	27	1	0	1	1	26	2	2
110	18	2	0	0	1	14	2	1
111	25	1	1	1	1	32	2	1
112	53	2	0	0	0	10	2	3
113	29	1	0	0	0	12	2	2
114	44	1	0	1	0	5	1	2
115	19	2	0	0	0	4	2	1
116	35	2	0	0	0	19	1	2
117	24	1	0	0	0	31	1	1
118	26	2	0	0	0	33	1	2
119	22	1	0	1	0	23	2	1
120	26	1	0	1	0	25	2	2
121	56	1	0	0	0	2	1	3
122	47	2	1	1	0	28	2	3
123	29	2	0	1	0	2	2	2
124	20	1	0	1	0	21	1	1
125	25	2	1	0	1	4	1	1
126	31	2	0	0	0	31	2	2
127	23	1	1	0	0	20	1	1
128	26	1	0	1	0	16	2	2
129	31	1	0	1	1	23	2	2
130	51	1	1	0	0	11	1	3
131	62	1	0	0	0	22	1	3

ID	age	sex	l_status	rest_use	dr_drink	durasi	status	kategori umur
132	52	2	0	0	0	26	1	3
133	33	1	0	0	1	7	1	2
134	40	1	0	0	0	3	1	2
135	27	2	0	0	0	16	1	2
136	39	1	0	0	0	13	1	2
137	57	1	0	1	0	5	2	3
138	28	1	0	1	0	20	2	2
139	20	1	0	1	0	33	2	1
140	30	1	1	1	0	6	2	2
141	24	2	0	0	0	15	2	1
142	92	2	0	0	0	31	2	4
143	25	1	0	1	0	24	1	1
144	42	1	0	1	1	10	1	2
145	63	1	1	0	0	24	1	3
146	66	1	0	1	0	24	2	4
147	26	1	0	0	1	9	2	2
148	27	1	0	1	1	17	2	2
149	69	2	0	0	0	2	2	4
150	48	2	0	0	0	10	1	3
151	42	2	1	0	0	30	1	2
152	28	2	0	0	0	9	1	2
153	65	1	0	0	0	28	1	4
154	36	1	0	0	0	24	1	2
155	50	2	0	0	0	6	1	3
156	19	1	0	0	0	25	1	1
157	23	1	0	0	1	3	1	1
158	23	1	1	1	1	29	1	1
159	72	2	0	0	0	24	2	4
160	36	1	1	0	0	31	1	2
161	24	1	0	0	1	30	1	1
162	27	2	0	1	0	10	2	2
163	45	2	0	0	0	24	1	3
164	28	1	1	1	1	24	2	2
165	50	1	0	0	0	18	1	3
166	63	2	0	0	0	17	2	3
167	19	2	0	0	0	17	1	1
168	25	2	0	0	0	15	1	1
169	27	2	0	0	0	15	1	2

ID	age	sex	l_status	rest_use	dr_drink	durasi	status	kategori umur
170	23	1	1	1	0	31	1	1
171	20	1	0	1	1	22	2	1
172	27	1	0	1	1	24	2	2
173	25	1	0	0	0	33	1	1
174	35	1	0	1	0	10	2	2
175	37	1	0	1	1	19	2	2
176	50	2	0	0	0	24	2	3
177	30	1	0	1	0	9	2	2
178	21	2	0	1	1	23	2	1
179	55	1	0	1	0	28	1	3
180	50	1	0	0	0	27	1	3
181	38	1	0	0	0	11	1	2
182	23	1	0	1	1	18	1	1
183	26	1	0	0	0	28	1	2
184	56	1	0	0	0	27	1	3
185	49	2	0	0	0	15	1	3
186	20	1	0	1	1	27	2	1
187	45	1	0	1	1	29	1	3
188	20	1	0	1	0	27	2	1
189	41	2	0	0	0	8	1	2
190	60	1	0	0	0	30	1	3
191	30	1	0	0	1	5	1	2
192	41	1	0	0	0	27	1	2
193	22	1	0	1	0	17	2	1
194	58	2	0	0	0	24	1	3
195	27	2	0	0	0	34	2	2
196	68	1	0	0	0	7	2	4
197	55	1	0	0	1	27	2	3
198	56	1	0	0	0	31	1	3
199	35	1	1	0	0	34	1	2
200	31	1	0	1	1	32	1	2
201	18	2	1	1	0	12	1	1
202	44	2	0	0	0	23	2	2
203	63	1	1	0	0	9	1	3
204	37	2	0	0	0	5	1	2
205	72	2	0	0	0	7	1	4
206	40	1	0	0	1	13	1	2
207	22	1	0	0	0	31	1	1

ID	age	sex	l_status	rest_use	dr_drink	durasi	status	kategori umur
208	24	1	0	1	0	14	2	1
209	76	1	0	1	1	10	2	4
210	53	1	1	1	0	29	1	3
211	25	1	0	1	1	24	2	1
212	66	1	0	1	0	16	1	4
213	46	1	1	1	1	23	1	3
214	27	1	0	0	0	20	1	2
215	22	1	1	0	1	35	1	1
216	49	2	0	1	0	25	1	3
217	58	1	0	0	0	26	1	3
218	25	1	0	0	0	16	1	1
219	43	1	1	1	1	6	2	2
220	58	1	0	1	0	12	2	3
221	22	2	1	0	0	15	1	1
222	46	1	0	0	0	20	1	3
223	46	1	0	0	0	23	1	3
224	40	2	1	1	0	24	2	2
225	36	1	1	1	0	14	2	2
226	83	1	0	1	0	28	2	4
227	21	1	0	1	1	30	2	1
228	52	2	0	0	1	15	2	3
229	53	2	0	0	0	30	1	3
230	20	2	1	0	1	29	1	1
231	24	1	0	1	0	30	2	1
232	51	2	0	1	0	31	2	3
233	31	1	1	0	1	21	1	2
234	22	1	0	1	1	5	2	1
235	32	2	0	1	0	8	1	2
236	62	2	0	0	1	29	2	3
237	18	1	0	1	0	17	1	1
238	54	1	0	1	0	23	1	3
239	45	1	0	1	1	24	2	3
240	48	2	1	1	1	5	1	3
241	38	1	1	1	0	24	2	2
242	21	1	0	1	1	18	2	1
243	77	2	0	1	0	23	2	4
244	16	1	0	0	0	28	1	1
245	25	1	0	0	0	24	1	1

ID	age	sex	l_status	rest_use	dr_drink	durasi	status	kategori umur
246	22	2	0	0	0	6	1	1
247	26	1	1	0	0	22	1	2
248	57	1	0	1	1	25	2	3
249	22	2	0	1	0	4	1	1
250	27	2	0	0	0	13	1	2
251	31	2	0	1	1	19	2	2
252	51	1	0	0	1	10	2	3
253	30	1	0	0	0	4	1	2
254	56	1	0	1	0	22	2	3
255	19	1	0	1	1	11	2	1
256	29	1	1	1	1	11	2	2
257	47	1	0	0	1	28	2	3
258	27	2	1	0	0	3	2	2
259	23	1	0	0	0	31	1	1
260	25	1	1	1	1	25	1	1
261	24	2	0	1	1	19	2	1
262	30	1	0	0	0	3	2	2
263	33	2	1	1	1	24	2	2
264	28	1	0	1	0	23	1	2
265	54	1	0	0	0	30	1	3
266	51	1	1	1	0	12	1	3
267	55	1	0	0	0	27	1	3
268	40	1	0	0	0	24	1	2
269	46	2	0	1	1	17	2	3
270	28	1	0	1	1	13	1	2
271	32	1	0	1	0	3	2	2
272	34	1	1	0	0	27	1	2
273	34	1	1	1	1	25	2	2
274	20	1	0	1	0	4	2	1
275	38	1	1	1	0	25	2	2
276	56	1	1	1	1	32	1	3
277	21	2	1	1	0	33	2	1
278	22	1	1	1	1	33	2	1
279	34	1	1	0	0	24	1	2
280	28	1	0	0	0	21	1	2
281	78	1	0	0	0	23	1	4
282	19	1	0	1	1	25	2	1
283	23	1	0	1	0	27	2	1

ID	age	sex	l_status	rest_use	dr_drink	durasi	status	kategori umur
284	38	1	0	1	1	16	2	2
285	31	1	0	1	0	24	2	2
286	40	2	0	1	0	11	2	2
287	27	1	0	1	0	8	2	2
288	20	2	0	0	0	24	1	1
289	24	1	0	0	0	29	1	1
290	45	2	0	0	0	27	2	3
291	22	1	0	1	0	30	2	1
292	28	1	1	1	0	28	2	2
293	48	1	0	0	0	17	2	3
294	23	1	1	1	1	28	2	1
295	37	1	1	1	0	7	2	2
296	56	2	0	1	0	22	2	3
297	58	2	0	0	1	28	2	3
298	45	1	0	0	1	11	1	3
299	23	1	0	1	1	33	2	1
300	35	2	0	0	0	9	1	2
301	23	2	0	1	1	25	1	1
302	21	2	0	1	1	9	1	1
303	70	1	0	1	0	17	2	4
304	71	1	0	0	0	20	2	4
305	46	1	0	1	0	18	2	3
306	57	1	0	1	0	23	2	3
307	21	1	0	1	0	21	2	1
308	22	1	1	1	1	14	1	1
309	21	1	0	0	1	20	1	1
310	52	1	1	0	0	33	1	3
311	42	2	0	1	0	29	2	2
312	30	1	1	1	1	2	2	2
313	48	1	0	1	1	23	2	3
314	31	1	0	0	1	2	2	2
315	28	1	0	0	0	20	1	2
316	32	1	0	0	1	3	1	2
317	46	1	1	1	1	26	2	3
318	24	1	0	0	0	2	1	1
319	54	1	0	1	0	15	2	3
320	25	1	0	1	0	25	2	1
321	36	1	1	0	0	22	1	2

ID	age	sex	l_status	rest_use	dr_drink	durasi	status	kategori umur
322	40	1	0	1	0	3	2	2
323	58	1	0	1	0	17	2	3
324	43	1	0	1	1	16	1	2
325	21	1	1	1	0	25	1	1
326	22	2	1	1	1	29	2	1
327	34	1	0	0	0	20	2	2
328	18	1	0	1	1	5	2	1
329	55	1	1	0	0	25	2	3
330	41	1	1	1	1	25	2	2
331	43	1	0	0	0	25	1	2
332	27	1	1	1	0	20	2	2
333	21	1	0	1	1	28	1	1
334	20	1	0	0	0	26	1	1
335	37	1	1	1	0	25	1	2
336	36	2	1	1	0	30	2	2
337	40	1	0	1	0	24	2	2
338	34	1	1	1	1	28	1	2
339	44	1	0	1	0	17	2	2
340	35	1	1	0	1	29	1	2
341	30	1	0	1	0	12	2	2
342	27	2	0	0	0	14	1	2
343	29	1	1	0	0	13	2	2
344	56	1	0	0	0	18	2	3
345	39	1	1	0	1	17	1	2
346	39	1	1	1	0	21	1	2
347	23	1	0	0	1	15	1	1
348	49	1	1	1	1	30	2	3
349	27	1	0	0	0	4	1	2
350	42	1	0	1	0	13	2	2
351	34	1	1	1	0	26	1	2
352	29	1	1	1	0	31	2	2
353	48	1	0	1	0	2	2	3
354	25	1	1	1	1	27	2	1
355	18	1	1	1	1	26	1	1
356	22	1	0	1	0	27	2	1
357	34	1	1	1	0	28	1	2
358	39	1	1	1	0	32	2	2
359	65	1	0	0	0	26	1	4

ID	age	sex	l_status	rest_use	dr_drink	durasi	status	kategori umur
360	29	2	1	0	1	33	1	2
361	23	1	1	0	1	25	1	1
362	45	1	0	1	1	6	2	3
363	37	1	1	1	1	30	2	2
364	29	1	1	1	0	25	2	2
365	33	1	0	0	0	6	2	2
366	21	1	0	1	0	17	1	1
367	58	1	1	0	0	13	2	3
368	39	1	1	0	0	24	1	2
369	26	1	1	1	0	34	1	2
370	29	2	1	0	1	26	1	2
371	37	1	1	1	1	28	2	2
372	29	1	0	0	0	28	1	2
373	80	1	0	0	1	6	1	4
374	44	1	0	1	1	21	2	2
375	32	1	0	0	0	5	1	2
376	33	2	1	0	1	11	1	2
377	17	2	0	0	0	22	1	1
378	21	1	0	1	1	32	2	1
379	22	1	0	1	0	34	2	1
380	34	1	0	0	0	29	1	2
381	35	2	0	0	0	4	1	2
382	30	1	0	0	0	28	1	2
383	65	1	0	1	0	19	2	4
384	35	2	0	1	1	28	2	2
385	60	1	0	0	0	16	1	3
386	32	1	1	1	1	20	2	2
387	45	1	0	1	1	13	2	3
388	24	1	0	1	1	17	2	1
389	21	2	0	1	0	2	2	1
390	57	2	1	1	0	4	2	3
391	58	1	0	0	0	22	1	3
392	22	1	0	1	1	9	1	1
393	56	1	0	1	0	31	2	3
394	34	1	0	1	0	11	1	2
395	49	1	0	1	0	10	2	3
396	27	1	0	1	0	12	1	2
397	51	1	0	1	0	31	1	3

ID	age	sex	l_status	rest_use	dr_drink	durasi	status	kategori umur
398	32	1	0	0	0	14	2	2

Lampiran 2. Output R Estimasi Parameter Model Cox

```
> cox < - coxph(Surv(durasi,status) ~ age + sex + l_status + rest_use + dr_drink,
  method="breslow", data=crash2)

> summary(cox)
Call:
coxph(formula = Surv(durasi, status) ~ age + sex + l_status + rest_use + dr_drink,
  data = crash2, method = "breslow")

  n= 398, number of events= 192

                coef       exp(coef)  se(coef)   z        Pr(>|z|)
age             0.012404   1.012481   0.004705   2.636    0.00838    **
sex             0.221632   1.248112   0.173730   1.276    0.20205
l_status       -0.536257   0.584934   0.166715  -3.217    0.00130    **
rest_use        1.131677   3.100851   0.169729   6.668    2.6e-11    ***
dr_drink        0.188945   1.207974   0.158201   1.194    0.23235
---
Signif. codes:  0 '***'    0.001 '**' 0.01 '*'    0.05 '.'  0.1 '        1
                exp(coef)  exp(-coef) lower .95  upper .95
age             1.0125     0.9877     1.0032     1.022
sex             1.2481     0.8012     0.8879     1.754
l_status        0.5849     1.7096     0.4219     0.811
rest_use        3.1009     0.3225     2.2233     4.325
dr_drink        1.2080     0.8278     0.8859     1.647

Concordance= 0.66  (se = 0.024 )
Rsquare= 0.153   (max possible= 0.993 )
Likelihood ratio test = 65.9  on 5 df,   p=7.281e-13
Wald test            = 61.26  on 5 df,   p=6.68e-12
Score (logrank) test = 65.53  on 5 df,   p=8.716e-13

> cox0 < - coxph(Surv(durasi,status) ~1, method="breslow", data=crash2)

> summary(cox0)
Call:  coxph(formula = Surv(durasi, status) ~ 1, data = crash2,  method =
  "breslow")

Null model
  log likelihood= -980.4182
  n= 398

> cox$loglik
[1] -980.4182 -947.4668
```

```
> cox0$loglik
[1] -980.4182

> 1-pchisq(-2*(cox0$loglik-cox$loglik[2]),5)
[1] 7.280843e-13
```

Lamaran 3. Output Pemilihan Model Terbaik Model Cox

```
> #variabel umur

> cox1 < - coxph(Surv(durasi,status) ~age, method="breslow", data=crash2)

> cox1$loglik
[1] -980.4182 -979.0736

> 1-pchisq(-2*(cox0$loglik-cox1$loglik[2]),1)
[1] 0.1010297

> #variabel jenis kelamin

> cox2 < - coxph(Surv(durasi,status) ~sex, method="breslow", data=crash2)

> cox2$loglik
[1] -980.4182 -980.4059

> 1-pchisq(-2*(cox0$loglik-cox2$loglik[2]),1)
[1] 0.8752442

> #variabel sim

> cox3 < - coxph(Surv(durasi,status) ~l_status, method="breslow", data=crash2)

> cox3$loglik
[1] -980.4182 -977.2083

> 1-pchisq(-2*(cox0$loglik-cox3$loglik[2]),1)
[1] 0.01128594

> #variabel sabuk pengaman

> cox4 < - coxph(Surv(durasi,status) ~rest_use, method="breslow", data=crash2)

> cox4$loglik
[1] -980.4182 -957.9863

> 1-pchisq(-2*(cox0$loglik-cox4$loglik[2]),1)
[1] 2.112355e-11


> #variabel alkohol

> cox5 < - coxph(Surv(durasi,status) ~dr_drink, method="breslow", data=crash2)
```

```
> cox5$loglik
[1] -980.4182 -978.5448

> 1-pchisq(-2*(cox0$loglik-cox5$loglik[2]),1)
[1] 0.05290849

> #variabel sabuk pengaman + umur

> cox41 < - coxph(Surv(durasi,status) ~rest_use + age, method="breslow", data=
  crash2)

> cox41$loglik
[1] -980.4182 -954.1608

> 1-pchisq(-2*(cox4$loglik[2]-cox41$loglik[2]),1)
[1] 0.00567372

> #variabel sabuk pengaman + jenis kelamin

> cox42 < - coxph(Surv(durasi,status) ~rest_use + sex, method="breslow", data=
  crash2)

> cox42$loglik
[1] -980.4182 -956.8910

> 1-pchisq(-2*(cox4$loglik[2]-cox42$loglik[2]),1)
[1] 0.1388515

> #variabel sabuk pengaman + sim

> cox43 < - coxph(Surv(durasi,status) ~rest_use + l_status, method="breslow",
  data= crash2)

> cox43$loglik
[1] -980.4182 -951.9841

> 1-pchisq(-2*(cox4$loglik[2]-cox43$loglik[2]),1)
[1] 0.000530758

> #variabel sabuk pengaman + alkohol

> cox45 < - coxph(Surv(durasi,status) ~rest_use + dr_drink, method="breslow",
  data=crash2)

> cox45$loglik
```

```
[1] -980.4182 -957.9672

> 1-pchisq(-2*(cox4$loglik[2]-cox45$loglik[2]),1)
[1] 0.8449337

> #variabel sabuk pengaman + sim + umur

> cox431 < - coxph(Surv(durasi,status) ~rest_use + l_status + age, method=
  "breslow", data=crash2)

> cox431$loglik
[1] -980.4182 -948.9244

> 1-pchisq(-2*(cox43$loglik[2]-cox431$loglik[2]),1)
[1] 0.01336971

> #variabel sabuk pengaman + sim + jenis kelamin

> cox432 < - coxph(Surv(durasi,status) ~rest_use + l_status + sex,
  method="breslow" , data=crash2)

> cox432$loglik
[1] -980.4182 -951.0706

> 1-pchisq(-2*(cox43$loglik[2]-cox432$loglik[2]),1)
[1] 0.1764818

> #variabel sabuk pengaman + sim + alkohol

> cox435 < - coxph(Surv(durasi,status) ~rest_use + l_status + dr_drink, method =
  "breslow", data=crash2)

> cox435$loglik
[1] -980.4182 -951.7158

> 1-pchisq(-2*(cox43$loglik[2]-cox435$loglik[2]),1)
[1] 0.4638452

> #variabel sabuk pengaman + sim + umur + jenis kelamin

> cox4312 < - coxph(Surv(durasi,status) ~rest_use + l_status + age + sex,
  method= "breslow" , data=crash2)

> cox4312$loglik
[1] -980.4182 -948.1732
```

```
> 1-pchisq(-2*(cox431$loglik[2]-cox4312$loglik[2]),1)
[1] 0.2203092

> #variabel sabuk pengaman + sim + umur + alkohol

> cox4315 < - coxph(Surv(durasi,status) ~rest_use + l_status + age + dr_drink,
  method="breslow", data=crash2)

> cox4315$loglik
[1] -980.4182 -948.2519

> 1-pchisq(-2*(cox431$loglik[2]-cox4315$loglik[2]),1)
[1] 0.2461638

> #variabel sabuk pengaman + sim + umur + jenis kelamin + alkohol

> cox43125 < - coxph(Surv(durasi,status) ~rest_use + l_status + age + sex +
  dr_drink ,method="breslow", data=crash2)

> cox43125$loglik
[1] -980.4182 -947.4668

> 1-pchisq(-2*(cox4312$loglik[2]-cox43125$loglik[2]),1)
[1] 0.2345826

> # Estimasi Parameter model cox terbaik dengan seleksi forward #

> cox431 <- coxph(Surv(durasi,status) ~rest_use + l_status + age,
  method="breslow", data=crash2)

> summary(cox431)
Call:
coxph(formula = Surv(durasi, status) ~ rest_use + l_status + age, data = crash2,
  method = "breslow")
  n= 398, number of events= 192
```

	coef	exp(coef)	se(coef)	z	Pr(>|z|)	
rest_use	1.134048	3.108213	0.162828	6.965	3.29e-12	***
l_status	-0.517437	0.596046	0.164976	-3.136	0.00171	**
age	0.011687	1.011755	0.004609	2.536	0.01122	*

Signif. codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'	0.1 '	1
	exp(coef)	exp(-coef)	lower .95	upper .95		
rest_use	3.108	0.3217	2.2590	4.2767		
l_status	0.596	1.6777	0.4314	0.8236		
age	1.012	0.9884	1.0027	1.0209		

```
Concordance= 0.664  (se = 0.024 )
Rsquare= 0.146   (max possible= 0.993 )
Likelihood ratio test   = 62.99  on 3 df,   p=1.351e-13
Wald test               = 58.78  on 3 df,   p=1.07e-12
Score (logrank) test    = 62.64  on 3 df,   p=1.602e-13

> #pemilihan model cox terbaik dengan interaksi

> #variabel sabuk pengaman + sim + umur + (umur*sim)

> coxint1 < - coxph(Surv(durasi,status) ~rest_use + l_status + age
  +(age*l_status), method="breslow", data=crash2)

> coxint1$loglik
[1] -980.4182 -948.7877

> 1-pchisq(-2*(cox431$loglik[2]-coxint1$loglik[2]),1)
[1] 0.6011169

> #variabel sabuk pengaman + sim + umur + (umur*sabuk pengaman)

> coxint2 < - coxph(Surv(durasi,status) ~rest_use + l_status + age
  +(age*rest_use), method="breslow", data=crash2)

> coxint2$loglik
[1] -980.4182 -948.3273

> 1-pchisq(-2*(cox431$loglik[2]-coxint2$loglik[2]),1)
[1] 0.2744999

> #variabel sabuk pengaman + sim + umur + (sim*sabuk pengaman)

> coxint3 < - coxph(Surv(durasi,status) ~rest_use + l_status + age + (l_status*
  rest_use ),  method="breslow", data=crash2)

> coxint3$loglik
[1] -980.4182 -948.8633

> 1-pchisq(-2*(cox431$loglik[2]-coxint3$loglik[2]),1)
[1] 0.7266332
```

Lampiran 4. Output Pengujian Parameter Model Cox

```
> #variabel umur

> cox1 < - coxph(Surv(durasi,status) ~age, method="breslow", data=crash2)

> summary(cox1)
Call:
coxph(formula = Surv(durasi, status) ~ age, data = crash2, method = "breslow")

  n= 398, number of events= 192

                coef        exp(coef)   se(coef)    z          Pr(>|z|)
age             0.007793    1.007824    0.004669     1.669     0.0951     .
---
Signif. codes:  0 '***'     0.001 '**'  0.01 '*'    0.05 '.'   0.1 '      1
                exp(coef)   exp(-coef)  lower .95   upper .95
age             1.008       0.9922      0.9986      1.017

Concordance= 0.537  (se = 0.024 )
Rsquare= 0.007   (max possible= 0.993 )
Likelihood ratio test = 2.69  on 1 df,   p=0.101
Wald test             = 2.79  on 1 df,   p=0.09509
Score (logrank) test  = 2.79  on 1 df,   p=0.09478

> #variabel sim

> cox3 < - coxph(Surv(durasi,status) ~l_status, method="breslow", data=crash2)

> summary(cox3)
Call:
coxph(formula = Surv(durasi, status) ~ l_status, data = crash2,   method =
  "breslow")

  n= 398, number of events= 192

                coef        exp(coef)   se(coef)    z          Pr(>|z|)
l_status        -0.4025     0.6687      0.1632      -2.467     0.0136     . *
---
Signif. codes:  0 '***'     0.001 '**'  0.01 '*'    0.05 '.'   0.1 '      1
                exp(coef)   exp(-coef)  lower .95   upper .95
l_status        0.6687      1.496       0.4857      0.9206

Concordance= 0.55  (se = 0.02 )
Rsquare= 0.016   (max possible= 0.993 )
Likelihood ratio test  = 6.42  on 1 df,   p=0.01129
```

Wald test = 6.08 on 1 df, p=0.01363
Score (logrank) test = 6.16 on 1 df, p=0.01305

> #variabel sabuk pengaman

> cox4 < - coxph(Surv(durasi,status) ~rest_use, method="breslow", data=crash2)

> summary(cox4)
Call:
coxph(formula = Surv(durasi, status) ~ rest_use, data = crash2, method =
"breslow")

n= 398, number of events= 192

	coef	exp(coef)	se(coef)	z	Pr(>|z|)	
rest_use	1.0170	2.7650	0.1606	6.332	2.42e-10	. ***

Signif. codes:	0 '***'	0.001 '**'	0.01 '*'	0.05 '.'	0.1 '	1

	exp(coef)	exp(-coef)	lower .95	upper .95
rest_use	2.765	0.3617	2.018	3.788

Concordance= 0.616 (se = 0.021)
Rsquare= 0.107 (max possible= 0.993)
Likelihood ratio test = 44.86 on 1 df, p=2.112e-11
Wald test = 40.09 on 1 df, p=2.423e-10
Score (logrank) test = 43.65 on 1 df, p=3.92e-11

Lampiran 5. Output Hasil Uji Asumsi Proportional Hazard

```
> ## plot log monus log survival ##
> #variabel kategori umur

> KMPLOT1< -survfit(Surv(durasi,status)~kategori.umur,data=crash2)
> plot(x=0,y=0, type="n",xlim=c(0,36), ylim=c(-5,2),xlab="waktu survival"
  ,ylab= "log(-log (survival))", main="plot log minus log survival variabel
  kategori umur")
> lines(KMPLOT1, fun
  ="cloglog",col=c("Red","Blue","Green","Yellow"),conf.int
=F,lty=c(1,1))
> legend(500,-2,c("1","2","3","4"),lty=2:3,col=c("Red","Blue","Green","Yellow")

> #variabel sim

> KMPLOT2< -survfit(Surv(durasi,status)~l_status,data=crash2)
> plot(x=0,y=0, type="n",xlim=c(0,36), ylim=c(-5,2),xlab="waktu
  survival",ylab= "log(-log (survival))", main="plot log minus log survival
  variabel sim")
> lines(KMPLOT2, fun="cloglog",col=c("Red","Blue"),conf.int=F,lty=c(1,1))
> legend(500,-2,c("Memiliki","Tidak"),lty=2:3,col=c("red","blue")

> #variabel sabuk pengaman

> KMPLOT3< -survfit(Surv(durasi,status)~rest_use,data=crash2)
> plot(x=0,y=0, type="n",xlim=c(0,36), ylim=c(-5,2),xlab="waktu
  survival",ylab= "log(-log (survival))", main="plot log minus log survival
  variabel sabuk pengaman")
> lines(KMPLOT3, fun="cloglog" ,col=c("Red","Blue"), conf.int=F, lty=c(1,1))
> legend(500,-2,c("Menggunakan","Tidak"),lty=2:3,col=c("red","blue")

> ## scaled schoenfeld residual ##
> cox431< -coxph(Surv(durasi,status)~rest_use+l_status+kategori.umur,
  methode="breslow",data=crash2)
> sresids < - residuals(cox431, type="scaledsch")
> colnames(sresids) < - names(cox431$coef)
> time < - as.numeric(rownames(sresids))
> plot(time, sresids[,1], xlab="Time", ylab="Scaled Schoenfeld Residual(sabuk
  pengaman)")
> lines(smooth.spline(time, sresids[,1]), col="red", lwd=2)
> plot(time, sresids[,2], xlab="Time", ylab="Scaled Schoenfeld Residual(sim)")
> lines(smooth.spline(time, sresids[,2]), col="red", lwd=2)
> plot(time, sresids[,3], xlab="Time", ylab="Scaled Schoenfeld
  Residual(kategori.umur)")
> lines(smooth.spline(time, sresids[,3]), col="red", lwd=2)
```